NBS
Publi-
cations

NBS TECHNICAL NOTE 1169

U.S. DEPARTMENT OF COMMERCE/National Bureau of Standards

Introduction to Noise in Solid State Devices

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress on March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, and the Institute for Computer Sciences and Technology.

THE NATIONAL MEASUREMENT LABORATORY provides the national system of physical and chemical and materials measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; conducts materials research leading to improved methods of measurement, standards, and data on the properties of materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

Absolute Physical Quantities[2] — Radiation Research — Chemical Physics — Analytical Chemistry — Materials Science

THE NATIONAL ENGINEERING LABORATORY provides technology and technical services to the public and private sectors to address national needs and to solve national problems; conducts research in engineering and applied science in support of these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the ultimate user. The Laboratory consists of the following centers:

Applied Mathematics — Electronics and Electrical Engineering[2] — Manufacturing Engineering — Building Technology — Fire Research — Chemical Engineering[2]

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides scientific and technical services to aid Federal agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following centers:

Programming Science and Technology — Computer Systems Engineering.

[1]Headquarters and Laboratories at Gaithersburg, MD, unless otherwise noted;
mailing address Washington, DC 20234.
[2]Some divisions within the center are located at Boulder, CO 80303.

Introduction to Noise in Solid State Devices

NBS technical note

Julius Cohen

Center for Radiation Research
National Measurement Laboratory
National Bureau of Standards
Washington, DC 20234

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, Secretary

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued December 1982

National Bureau of Standards Technical Note 1169
Natl. Bur. Stand. (U.S.), Tech. Note 1169, 63 pages (Dec. 1982)
CODEN: NBTNAE

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1982

For Sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
Price $4.75
(Add 25 percent for other than U.S. mailing.)

PREFACE

Noise is the expression of fluctuations in the numbers or velocities of an ensemble of mobile electrons. Although it may be a fundamental physical phenomenon sometimes beneficiently exploited, as in noise thermometry, it is its harmful aspects which account for so much activity aimed at its reduction. In electronics, for example, noise establishes the lower limit of a signal which can be detected or amplified; in radiometry, noise sets the lower limit of radiative power which can be detected with a sensor and adversely affects the accuracy of measurements.

Its importance having been long recognized, there is a considerable body of published literature dealing with noise. Notwithstanding, the theoretical reviews are by nature quite mathematical and tend to be esoteric and abstract; the qualitative reviews generally are more assertive than explanatory.

This is a didactic monograph intended to be readable by physicists and electronic engineers with only general background, and should be of particular value to beginners of applied work in the field of noise. Others who wish to gain an understanding of this important phenomenon also will profit. The attempt has been made here to simplify a complex subject, and the text is mainly qualitative. An aim is to impart a "feel for the subject", and while the treatment generally is not rigorous, it is not misleading. Although noise reduction as an ultimate goal is a main concern, the premise is that a basic understanding of noise processes forms a necessary base for the applications engineer or scientist as well as for the device designer.

This monograph is divided into two parts. The first, Theory, deals in detail with each of the four principal noises found in solid state devices. The emphasis here is on phenomenological aspects of noise, its origins, and its manifestations.

The second part, Applications, is a systematic treatment of noise in selected solid state devices. Analyses progress from a single noise source in a circuit element to four noise sources in a device; concomitantly equivalent circuits are developed to facilitate the solution of various complex noise problems. The devices considered here were chosen as much for illustrative value as for commercial importance or currentness. They reveal the evolutionary nature of electron devices and the transfer of concepts from the earlier to the later devices, thus providing a simplifying coherence. Because of the way this work is structured, it should be read successively.

Thanks are due Jon Geist for suggesting that I write such a monograph and for his continued support. I appreciate also Edward Zalewski's scrutinous review of the manuscript and helpful comments.

CONTENTS

Page

List of Figures

INTRODUCTION TO NOISE IN SOLID STATE DEVICES

Julius Cohen

Radiometric Physics Division
Center for Radiation Research

This is a short didactic monograph on electronic noise which aims to impart a "feel for the subject". The work is divided into two parts. The first, Theory, deals in detail with the principal noises found in solid state devices, namely shot noise, thermal noise, 1/f noise, and generation-recombination noise. The second part, Applications, is a systematic treatment of noise in selected solid state devices. Analyses progress from a single noise source in a circuit element to four noises in a device; concomittantly equivalent circuits are developed to facilitate the solution of various complex noise problems. Examples treated in this part include resistors, rectifiers, transistors, and photo-detectors. The work concludes with a recapitulation and useful references.

Key words: electronics; noise; photon detector; rectifier; solid state devices; transistor.

Part I. THEORY

1. Elementary Statistical Considerations

1.1 Terms

A few statistical terms first need to be defined:

A deviation is the difference between the arithmetic mean of a series of statistical data and any single datum belonging to the series.

Fluctuations consist of deviations, and connote irregularities, an uncertain shifting back and forth; e.g., temporal or spatial randomness of a physical entity or attribute. Thus, fluctuations are irregular, or non-systematic, deviations. A vivid example of fluctuations would be Brownian movement, or motion (to be described in sec. 3.2.1).

When dealing qualitatively with fluctuation phenomena involving large numbers of events, of which noise is an example, it is convenient to express data collectively in terms of a mean and some mean deviation from the mean. But the mean deviations (the arithmetic sum of the individual deviations in the series divided by the total number of deviations or data belonging to the series) is nearly zero, which per se carries little useful information and none whatsoever as regards the average magnitude of the deviations.

On the other hand, the mean square deviation, e.g., $\overline{(y-\bar{y})^2}$ or alternatively, the root mean square deviation, e.g., $\left(\overline{(y-\bar{y})^2}\right)^{\frac{1}{2}}$, does carry the desired useful

information, and therefore is used extensively in expressing fluctuations. To express the degree of fluctuation (i.e., the percentage of fluctuation about the mean value) one need only divide the mean square deviation by the square of the mean, e.g., $\left(\overline{(y-\bar{y})^2}\right)^{\frac{1}{2}}/(\bar{y}^2)^2$.

1.2 The One-dimensional Random Walk

The following example, the one-dimensional random walk, should help to make these statistical concepts more concrete. Picture a straight line with centerpoint x=0 and let the opposite directions of the line with respect to the centerpoint be designated the positive and the negative directions. Assume a particle, constrained to the line, to be located at the centerpoint at time t=0, and thereafter to be struck repeatedly with a constant impact such that each impact moves the particle randomly as to direction along the line, a fixed distance $\pm l$. Then,

$$l_n = \pm l, \tag{1}$$

where l_n denotes the n-th step.

Let x denote the <u>total</u> displacement after N steps. Then,

$$x = \sum_{n=1}^{N} l_n. \tag{2}$$

The symbol < > will be used to signify the <u>average</u> taken over an infinitely large group of many similar observations of the displacement x. Then

$$\langle x \rangle = \sum_{n=1}^{N} \langle l_n \rangle = 0. \tag{3}$$

Now consider instead, the quantity, x^2.

$$x^2 = (l_1 + l_2 + \cdots + l_N)\,(l_1 + l_2 + \cdots + l_N) = \sum_{n=1}^{N} l_n{}^2 + \sum_{n \neq m} l_n l_m. \tag{4}$$

$$\langle x^2 \rangle = \sum_{n=1}^{N} \langle l_n^{\,2} \rangle + \sum_{n \neq m} \langle l_n l_m \rangle. \tag{5}$$

But,

$$\sum_{n=1}^{N} \langle l_n^{\,2} \rangle = N\, l^2, \tag{6}$$

and

$$\sum_{n \neq m} \langle l_n\, l_m \rangle = 0^{*}. \tag{7}$$

Therefore,

$$\langle x^2 \rangle = N\, l^2, \tag{8}$$

which is an important fundamental equation relating the average magnitude of displacement with the total number of individual random events.

2. Designations of Noise

Although there are but few fundamental sources of noise, there is a large variety of noise processes and manifestations, and in this respect many particular kinds of noises have been encountered. Their designations have stemmed, not from systematics, but from natural linguistic processes; thus, particular noises may be named for a person, a source, a process, or a manifestation.

The lack of a taxonomy[1] can obscure identity or relationship, as when a single noise has multiple designations, or a noise is a derivative, extension, or special case of a more general noise. It is not the purview of this report to treat the large variety of noises which has been reported; instead, this report is confined to the common noises, specifically those which are prominent in solid state devices. For these noises, where alternative designations are in use, they will be indicated; also relationships will be given.

---

*Because each step is random as to direction, these products will occur with equal probability. Hence $\langle l_n\, l_m \rangle = 0$.

[1]An attempt to classify noise was made recently by van Vliet; see ref. 1, p 3.

3. The Four Principal Noises in Solid State Devices

The four principal types of noise in solid state devices are shot noise, thermal noise, 1/f noise, and generation-recombination noise. They are explained below.

3.1 Shot Noise

Shot[1] noise and thermal noise constitute the two most fundamental noises and are the progenitors of most other noises. Both are intimately bound with the structure and nature of matter: the former arises from the fact that electrons are discrete particles (governed by statistical laws); the latter is a kinetics phenomenon involving collisions between the free electrons of a substance and its lattice. Shot noise will be treated first because of its historical precedence.

Vacuum tubes and solid state devices are more closely related than is generally appreciated, and although shot noise was initially observed over 60 years ago in a thermionic vacuum diode, it is a common occurrence in modern solid state devices. To gain an understanding of why shot noise exists, where it should be expected to be found, and how it might be modified, it is instructive to consider the operation of a simple vacuum tube as it relates to shot noise.

3.1.1 Operation of a Vacuum Diode

A vacuum diode consists of two elements in an evacuated envelope: the cathode, the source of electrons; and the anode, the receiver, or collector of electrons. The cathode is the "heart" of the tube; sometimes it is a metal, usually it is a semiconductor. At the interface between cathode and vacuum there exists a potential barrier which may be thought of as a force which keeps the electrons from otherwise spilling out of the cathode. Electrons may escape from the cathode in one of two ways: by going over the barrier or by tunneling through the barrier, and both processes require energy to be put into the cathode. For the former, either heat (thermionic emission) or light (photoelectric emission) may be used; for the latter, an intense electric field (field emission) between cathode and anode may be applied which reduces barrier width (and height).

3.1.1.1 Thermionic Emission

The case of thermionic emission, wherein the cathode is heated, will be treated: Electrons are discrete particles, the collective properties of which

---

[1]The discoverer, W. Schottky, likened this noise to the audible effect made by an impinging stream of shot (schrot, in German).

are described by statistical laws. Thus, their velocity (and energy) distribution in a solid normally is given by Fermi-Dirac statistics for a metal, while classical Maxwellian statistics may be used for a semiconductor [2][1]. Heating the cathode liberates electrons internally and imparts energy to free electrons through the process of random atomic collisions by the lattice, and only those electrons which reach the vacuum interface with energy component normal to the cathode surface sufficiently high to surmount the barrier, will emerge. The emission is random in time as it is inconceivable that equally large numbers of electrons could reach the surface possessing the required energy and direction, in equally minute increments of time. And regardless of what the form of the electron distribution was inside of the cathode, the electrons emerging into free space will assume a Maxwellian distribution of velocities [3].

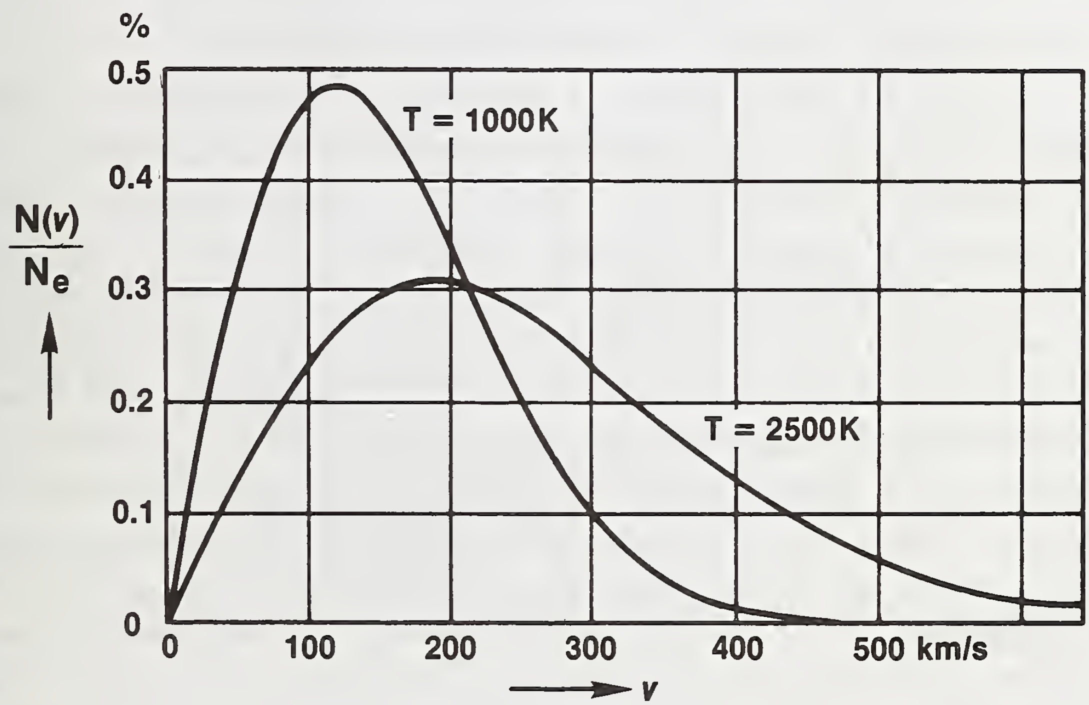

Figure 1. Maxwellian velocity distribution of the emitted electrons (Herrmann and Wagener [4]).

Figure 1 shows plots of the Maxwelliam velocity distribution of the emitted electrons for two different cathode temperatures; N(v) is the number of emitted electrons having velocity v, and N_e is the total number of emitted electrons.

[1]Figures in brackets indicate the literature references at the end of this paper.

Thus, the ordinate gives the fraction of emitted electrons possessing a velocity corresponding to v, or alternatively, the probability[1] of an emitted electron having any particular velocity. The curves show that the probability falls somewhat rapidly either side of the maximum probability, a characteristic which becomes more pronounced the lower the temperature. Hence, the totality of emitted electrons may be treated approximately as being endowed with the most probable velocity. In addition, the fact that the emitted electrons obey Maxwell's distribution law constitutes a proof that they are free; i.e., each emitted electron is independent of the others.

So far, only the emission of electrons from the cathode has been considered, above. To obtain an emission current, however, requires that the electrons be collected at the anode, and the role of the anode in determining emission current characteristics will be considered next; after which the shot noise may be treated.

3.1.1.2 The Three Ranges of Emission Current Characteristics

First, in order to obtain an emission current, a voltage is required to be applied to the anode. This determines the distribution of potential in the interspace, thereby influencing the collection of the emitted electrons. Thus, the operating conditions of a diode are cathode temperature, which (for a given cathode material) establishes the rate of electron emission, and anode voltage, which determines the fraction of emitted electrons collected. The operating conditions determine the operating characteristics (of the emission current) of the diode, and three different regions of characteristics may be distinguished, as shown in figure 2: (1) Saturation, or temperature-limited region -- If an adequate positive voltage is applied to the anode, all of the electrons emerging from the cathode surface will reach the anode. As shown in the figure, the emission current at fixed temperature increases slowly with increasing anode voltage; i.e., saturates, and the magnitude of the emission current is essentially determined by the cathode temperature. (2) Space-charge-limited region -- If the anode voltage is not sufficient for saturation, only a fraction of the emitted electrons can reach the anode, with the remainder forming a negative space charge in front of the cathode. This causes the potential barrier in the interspace to increase; the less energetically emitted electrons then cannot surmount the barrier and fall back to the cathode, thus reducing the emission current. The curves of figure 2 show the space-charge-limited currents to be quite sensitive to anode voltage.

---

[1]Maxwell's law was originally deduced purely from considerations of probability; see ref. 5.

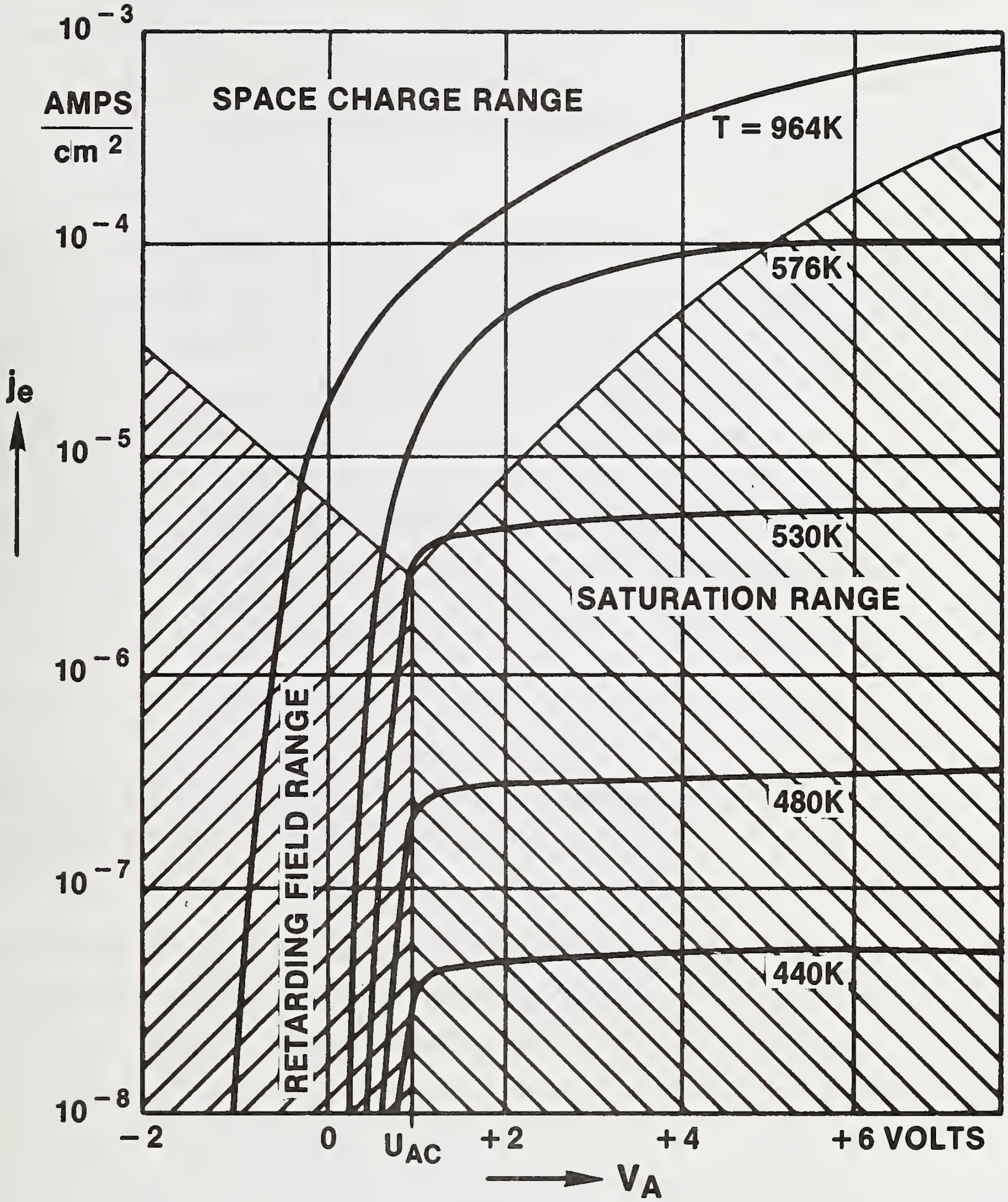

Figure 2. The three ranges of the characteristics. Electron densities j_e as a function of anode voltage for different temperatures of the cathode; cathode work function $\sim$ 1.4 eV (Herrmann and Wagener [4]).

(3) Retarding field region -- If the anode voltage is negative or only weakly positive, a substantial potential barrier is presented to the emitted electrons, of which only the very energetic can reach the anode, and the emission current is drastically reduced. This region is mentioned only for completeness; it is of little practical importance and need not be discussed further.

3.1.2 Full Shot Noise

As concerns noise, the saturated region is of particular significance because only in this region can "pure", or full shot noise be obtained. The reason, alluded to above, is that only in this region do the emitted electrons behave as totally free, random particles. In other words, only the saturated current is comprised of electrons which have been randomly emitted in time and which remain free, from emergence from the cathode to collection at the anode. For an assembly of particles so described, a well-known theorem in statistics is applicable, which states that the mean square of the fluctuations in a number of independent, random particles is equal to the mean number of particles; i.e.,

$$\langle (n - N)^2 \rangle = N, \tag{9}$$

where n is the number of electrons emitted in a time interval t, and N is the mean value of n. (Note the similarity of this equation with that for the 1-dimensional random walk, previously derived.)

Fluctuations in current are equivalent to noise. Current, by definition, is proportional to the number of electrons and the (transit) velocity; however, as argued earlier, all of the emitted electrons may be assumed to have approximately the same velocity. Therefore, the fluctuations in current are essentially fluctuations only in the numbers of electrons, and the source of the noise is the random emission of individual electrons from the cathode. The electrons, emitted from the cathode in pulses, are collected at the anode in pulses. Further, from the above statistical theorem, the fluctuations in current--hence, the noise--should be proportional to the square root of the average, or steady, current.

An equation for calculating the shot current/noise can be found by expanding the current pulse associated with the passage of an electron into a Fourier series, and adding the terms of this series lying in a certain bandwidth[1]. Schottky [6], examining the noise in the saturated region, derived the following formula for the

[1]For a detailed derivation of shot noise, see section B of the Appendix.

mean shot current within a bandwidth Δf (equal to the frequency band f to f+Δf, and during measurement determined by the bandwidth of the amplifier):

$$\overline{i}_{sh} = (2eI_s\Delta f)^{\frac{1}{2}}, \tag{10}$$

where I_s is the saturated current of the diode (also, the steady, average emission current in the saturated region).

Thus, the shot current is independent of the frequency[1] (as expected, because of its random nature), and, as expected, determined only by the magnitude of the saturation current and the bandwidth. The dependence of the shot noise on the square root of the current is a unique characteristic.

3.1.3 Partial Shot Noise

The validity of Schottky's equation has been experimentally verified and is well established [7]. Nevertheless, vacuum tubes are seldom used as temperature-limited current devices; the space-charge-limited mode is normal. When a diode is operated in the space-charge-limited mode, it exhibits much lower noise [7] than it would if it were operated in the temperature-limited mode passing the same emission current[2]; as in shot noise, the noise current is frequency independent. The mechanism whereby space charge causes reduction in noise is not well understood; a possible explanation which is rather involved is given elsewhere [7].

3.1.4 Equivalent Circuits: Shot Noise and Partial Shot Noise

Finally, some equivalent circuits for a diode with temperature-limited current (shot noise) and a diode with space-charge-limited current (reduced shot noise), respectively, are given in figure 3 and figure 4 (a) and (b).

3.2 Thermal Noise (a.k.a. Johnson noise; Brownian noise)

3.2.1 Brownian Movement in Fluids

In 1827, the English botanist, Robert Brown [8], observed a phenomenon which was to have great significance. Although Brown initially looked at microspores (pollen) in water, the phenomenon can be readily observed if any small particles suspended in a liquid are examined with a powerful microscope. The particles are seen to be darting about to and fro, spontaneously, randomly, and continuously. Each particle spins, rises, sinks, and rises again. This irregular

---

[1]Frequency-independent noise is commonly termed white noise.

[2]Strictly, shot noise refers to the noise accompanying temperature-limited, or saturated, emission current; however, the reduced noise is referred to as "partial shot noise" and "reduced shot noise".

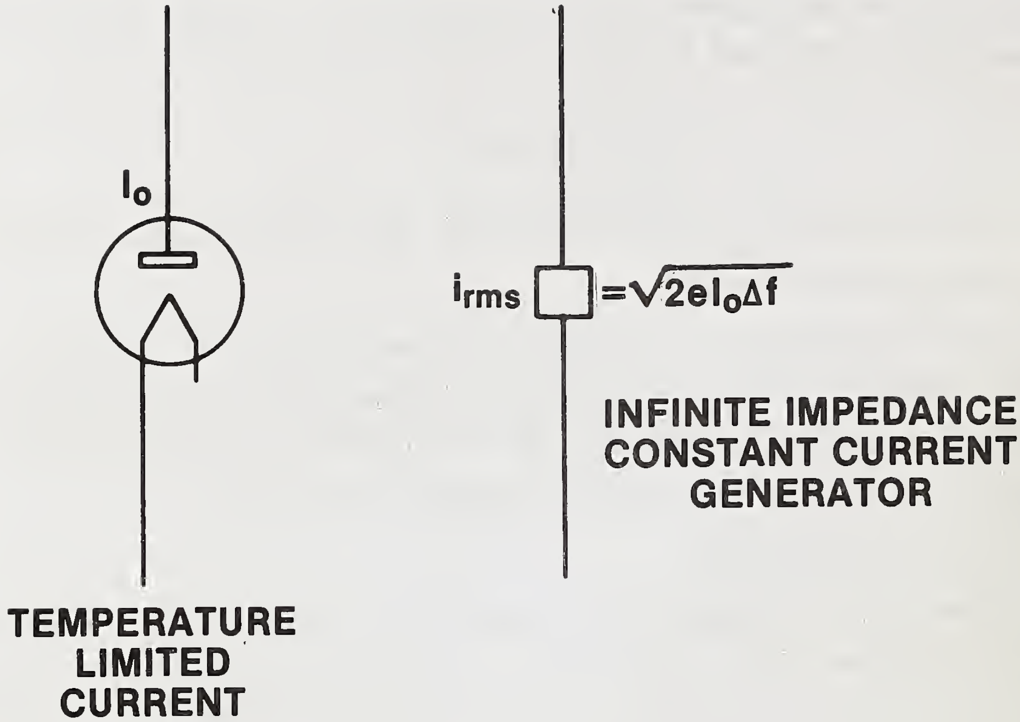

Figure 3. Constant-current generator equivalent of a diode with temperature-limited current (Spangenberg [7]).

motion is referred to as the Brownian movement/motion of the particle. It is independent of the nature, or source, of the suspended particles, all particles of the same size being equally in motion; smaller particles, however, move more vigorously than larger ones.

A valid explanation for the phenomenon was not forthcoming for nearly eight decades after its discovery, in spite of the many attempts made by various investigators. Indeed, Brown himself misinterpreted the origin of the phenomenon he had discovered. It was Einstein [9] who proved in a series of classical papers that Brownian movement arose directly from the incessant and random bombardment of the particles by the molecules of the surrounding fluid. Thus, the phenomenon of Brownian movement is a direct proof of the existence of random molecular motion, and thereby of the discrete molecular structure of matter.

3.2.2 Einstein's Prediction of "Brownian Noise" in Solids

Einstein proceeded to argue that an analagous Brownian movement must exist inside a solid by virtue of collisions between the free electron particles of the solid and the atoms comprising its lattice. Two further assumptions were made:

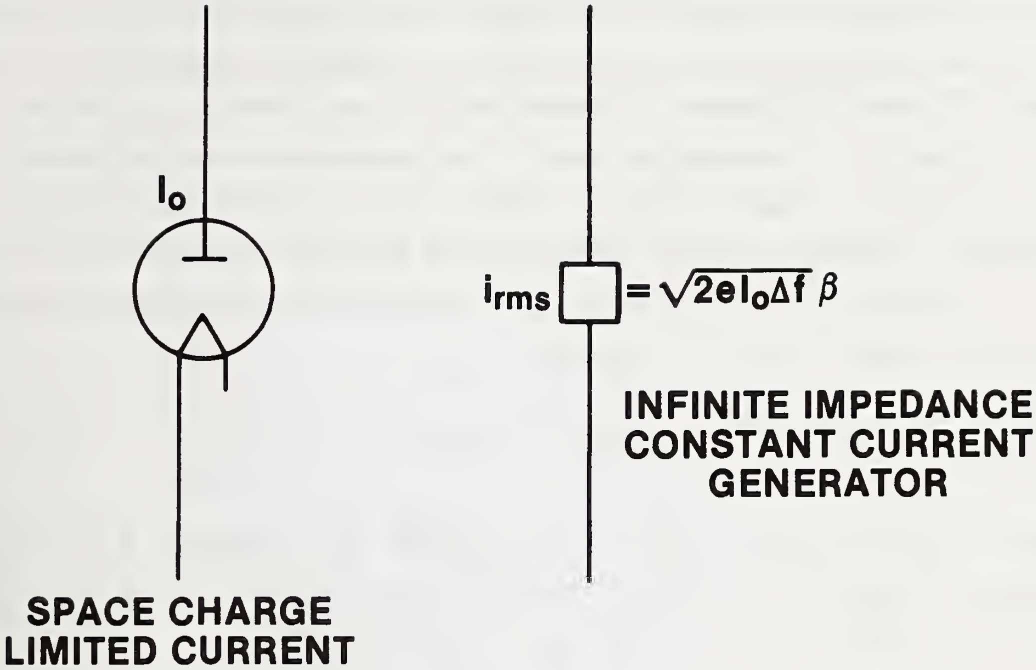

Figure 4(a). Constant-current generator equivalent of a diode with space-charge-limited current; $0 > \beta << 1$.

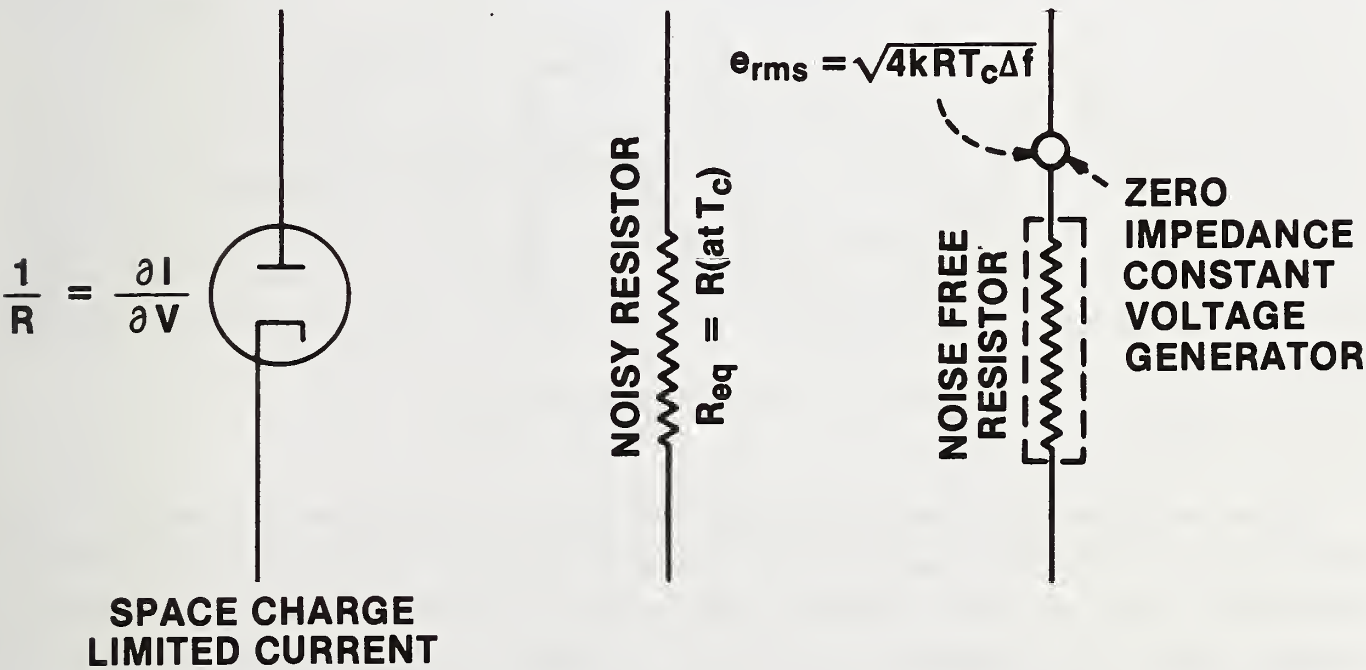

Figure 4(b). Resistor equivalent to a diode with space-charge-limited current (Spangenberg [7]).

1) that the free electrons behaved like a perfect gas--thus investing otherwise sterile particles with pressure or force on the one hand and temperature or energy on the other; 2) that the degree of thermal agitation of the electrons was influenced by a "viscosity" of the medium (here, its conductance--the reciprocal of resistance). Then, by extension of his random-walk analysis for the Brownian motion in liquids, Einstein derived the following equation for the fluctuations in charge δq, in a material of conductance G, during an interval t, thus predicting the existence of thermal noise in a solid:

$$\langle \delta q^2 \rangle = 2kTGt. \tag{11}$$

Although this equation is not in a practicable form, owing to a time rather than a frequency interval, it is nevertheless correct (and is consistent with the usual and useful form that was derived years later by others).

Brownian movement is spontaneous and eternal; so too, thermal noise. It may be reduced, but short of destroying the matter which it infests, it cannot be eliminated.

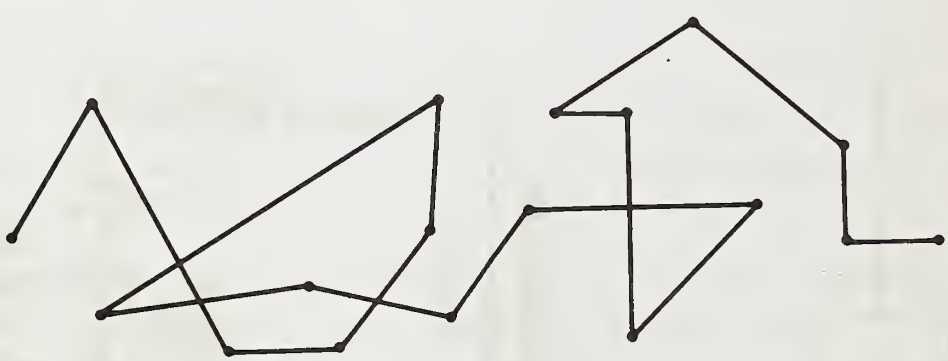

Figure 5. Illustration of free path.

One can visualize--simply, yet physically justifiably--hordes of miniscule particles, free electrons suspended in interstitial space, falling into collosal, sluggishly vibrating atoms from which they bounce off wildly in any and all directions, repeating the process, ad infinitum. A schematic diagram of the Brownian movement of a single particle is depicted in figure 5. The path traversed between two successive collisions is a straight line (described with constant velocity), and the course consists of a series of zigzag paths of various lengths. The average length of a path is the mean free path, defined as the sum of the lengths

of a large number of paths divided by the total number. (This quantity is of great importance in transport phenomena, in general.) Just as in the 1-dimensional random walk, the result after many collisions will be that there is no net current--each motion by an electron being compensated by an equivalent opposite motion by another electron--but there is a fluctuating current, attributable primarily to the paths or directional changes in velocity, not to the number of free electrons, which remains essentially constant.

3.2.3 Johnson's Discovery of Thermal Noise

Some 20 years after Einstein's prediction of thermal noise--perhaps detection was not technologically feasible earlier--J. B. Johnson [10], working at The Bell Telephone Laboratories, measured the thermal noise in a wide variety of conductors, and determined empirically that the mean-square potential fluctuation[1] across the conductor is proportional to the electrical resistance (the inverse of conductance) and the absolute temperature of the conductor; it is independent of the size, shape, and material of the conductor. These findings are in full agreement with Einstein's prediction: the "viscosity" of the conductor shows up here as resistance, the inverse of conductance, because fluctuations in voltage, not in current, are described; and just as the nature or origin of the particles was immaterial in their Brownian movement in fluids, so where the electrons came from is irrelevant here.

3.2.4 Nyquist's Formula

Diverse disciplines such as thermodynamics, statistics, and statistical mechanics have each been used to derive the (same) equation for thermal noise [11,12,13]. Nyquist's [11] pithy deduction of the formula using thermodynamical reasoning is, however, renowned for its elegance[2]. The well-known expression for thermal voltage noise in a conductor of pure resistance R and of temperature T is

$$\overline{v^2} = 4kTR\Delta f, \qquad (12)$$

where $\overline{v^2}$ is the mean square of the voltage fluctuation within the frequency range Δf and k is Boltzmann's constant.

---

[1]Actually, noise phenomena originate as fluctuations in current, not voltage; however, it is more convenient to speak of a disturbance to an amplifier as a voltage fluctuation at the input rather than as a current fluctuation at the output.

[2]For a detailed derivation of thermal noise, after Nyquist, see section A of the Appendix.

It will be noted that such quantities as charge, number, and velocity of electrons do not appear explicitly in the formula for voltage or noise; they are, however, implicit in R.

The above noise formula refers to the open-circuit voltage across the resistor R; alternatively (by Ohm's law), the noise may be expressed as the short-circuit current through the resistor R:

$$\overline{i^2} = 4kTG\Delta f. \tag{13}$$

3.2.5 Two Paradoxes

It is instructive to consider the resolution of two paradoxes relating to measurements in thermal noise. The first, with metaphysical overtones: If, indeed, there is no net current in a passive conductor in thermal equilibrium and the amount of current drawn by the measurement instrument is negligible, how is the noise current/voltage able to be measured? The second paradox: The formula for thermal noise is derived for a passive conductor in thermal equilibrium. Why doesn't the passage of a steady (small) current through a resistance increase the fluctuation voltage (it remains the same)?

The answer to the first question is to be found in Nyquist's theorem [11]; it is because thermodynamically there must be a transfer of power from the "noisy" resistor to the input terminal of the instrument, considered initially noiseless. Alternatively, from a circuitry viewpoint, the 'noisy' source resistor may be thought of as becoming incorporated into the circuitry of the measurement instrument.

The answer to the second question was given by Bell [14], and is restated here. It is that within a (solid) conductor, collisions between free electrons and atoms are extremely frequent, and the drift velocity acquired under an electric field is small compared with the thermal agitation velocity. Thus, the number of random events, fluctuations, is insensitive to the average current passing through the conductor. The effect of passing a current may be likened to walking while holding a screen on which a motion picture is being shown; the movie doesn't change, it merely translates.

3.2.6 Comparison of Thermal- and Shot Noises

Thermal noise and shot noise have been seen to have commonalities as well as significant differences. On the one hand, both are of fundamental nature, derive from the random events of free electrons, and (because of the randomness)

are frequency independent. On the other, thermal noise occurs in a passive (or active) resistor and is unaffected by passage of a steady current; shot noise occurs only if a current is passed, and the noise is proportional to the square root of the steady current. Thermal noise is thermal agitation noise, originating in collisions between free electrons and lattice atoms at temperature T which produce fluctuations in the velocity of free electrons; shot noise originates in the fluctuations in the number of free electrons. Notwithstanding the contrary assertions of early investigators [15], thermal noise and shot noise are two distinctly different noises (and attempts to unify them result in physically unpalatable interpretations).

3.2.7 Equivalent Circuits of a Noisy Resistor

The equivalent circuits of a resistor with thermal noise, of which there are two, are shown in figure 6. As shown, the effect of thermal noise in a resistor can be expressed either as an emf in series with the resistor, considered noiseless, or as a constant-current generator in parallel with the resistor, considered noiseless.

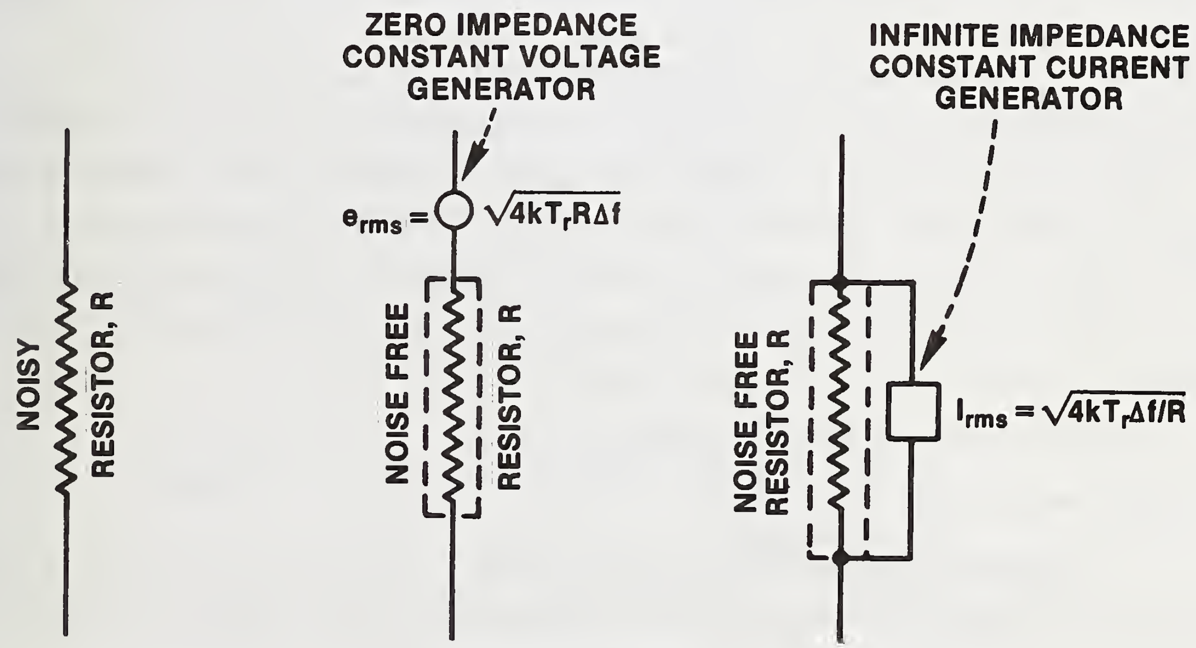

Figure 6. Equivalent circuits of a noisy resistor (Spangenberg [7]).

3.3 1/f Noise (a.k.a. flicker noise)

The distinguishing characteristic of 1/f (one-over-f) noise is its frequency response; the square of the noise current is approximately proportional to the inverse of electrical frequency. It is also approximately proportional to the square of the steady current. These fluctuations were discovered by Johnson [16] while studying vacuum tubes, and have been investigated in detail by others [17,18]. If (atomic film- and oxide) cathodes are examined at low frequencies (< 10 kHz), fluctuations are observed which are (often considerably) larger than that given by the shot-current equation, and as for shot noise from vacuum diodes, 1/f noise is lower in the space-charge current region than in the saturated current region. Further, experiments suggest that 1/f noise increases with increasing temperature. Johnson attributed the 1/f noise to fluctuations in cathode work function.

Unlike thermal noise and shot noise which are fundamental and intrinsic, 1/f noise is apparently extrinsic, being particularly sensitive to surface phenomena and imperfections. Thus, thin film- and polycrystalline devices are especially susceptible. The unwanted phenomenon is commonly encountered in widely used solid state devices such as photodiodes and the various mutants of field-effect transistors, and is being intensively investigated. Previously, 1/f noise had been observed in many other devices such as metal and carbon resistors, crystal rectifiers, vacuum tubes, and photoconductors[1]. The phenomenon has been attributed variously, as appropriate, to variation in contact potential, ionic conduction/migration, instabilities of (metal point) contacts, surface states/conduction, tunneling into traps, dislocations, diffusion of defects, defect-defect scattering, and fluctuations in mobility. Notwithstanding, the mechanisms--apparently there are more than one--have not yet been adequately explained.

3.4 Generation-Recombination Noise

3.4.1 Mechanisms

Generation-recombination noise, commonly referred to as g-r noise[2] is caused by fluctuations in the generation, recombination, and trapping of carriers

---

[1]As an example of the universality of 1/f phenomena, T. Musha has shown that a [Japanese] man standing on only one leg generates a 1/f spectrum; see ref. 1, p 143.

[2]Sometimes, especially in the early literature, g-r noise is referred to as shot noise.

in homogeneous semiconductors (and insulators), particularly photoconductors, which in turn cause the number of carriers to fluctuate. Hence, the conductance of the sample fluctuates. The mechanism underlying these processes is the quantum-mechanical transitions of electrons between designated energy states of the conduction band, the valence band, localized impurity levels, traps, or recombination centers. The transitions of concern in g-r noise are specifically transitions between the conduction band and the valence band, and between impurity levels (or traps, or recombination centers) and the conduction- or valence band. The electrons making transitions spend various amounts of time in each state they occupy, according to the designation of the state (i.e., the conduction band, the valence band, etc.) and the number of states per energy level, and the occupancy of the states depends on the particular system studied.

G-r noise may be regarded as a generalized form of thermal noise in which the Fermi level is caused to fluctuate. In g-r noise, however, the conductance of the solid fluctuates because of the fluctuations in the number of carriers, while in thermal noise the number of carriers is essentially constant, and the noise stems from fluctuations in the velocity of the carriers.

3.4.1.1 The Two-level System

Formulas for calculating g-r noise are usually derived with the use of probability statistics, and the simplest system to consider is the two-level system in a semiconductor where carrier transitions can occur only between two energy levels, as previously specified; e.g., between the valence band and the conduction band or between the conduction band and a set of impurity levels. This case is depicted schematically in figure 7. Here E and E' denote two energy

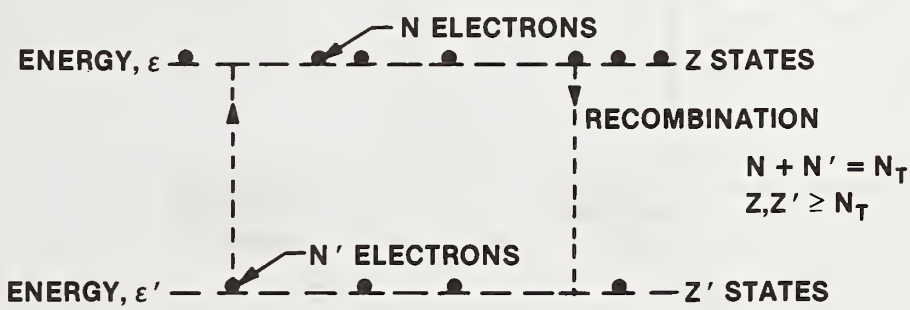

Figure 7. Two-level g-r problem (van Vliet and Fassett [19]).

levels containing respectively Z and Z' possible states; the numbers of electrons occupying these states are denoted respectively by N and N', their sum N_t being constant. Only one electron is allowed per state (in accordance with the Pauli exclusion principle), and not every state is occupied by an electron. A transition from E' to E is a generation, and that from E to E' is a recombination.

The probability of a transition is usually assumed to be proportional to the product of the number of electrons available to make the transition and the number of unoccupied states in the other level which can accept the electrons. Thus, expressions for the generation- and recombination probabilities can be derived, and an expression for the variance in the number of carriers $\langle \Delta N^2 \rangle$ obtained, as well as noise expressions for particular cases [19].

The solutions are usually of academic interest, however, because of the apparent lack of experimental confirmation for the existence of the two-level model[1]. Indeed, many experimental studies [20] of g-r noise in various systems suggest that either the two-level model is an unrealistic simplification or the noise phenomena are obscured by other processes; e.g., space charge, impact ionization, or surface recombination.

3.4.1.2 The Three-level System

It is desirable therefore to consider a semiconductor system wherein electronic transitions take place between three or more energy levels, and two examples are depicted schematically in figures 8(a) and (b).

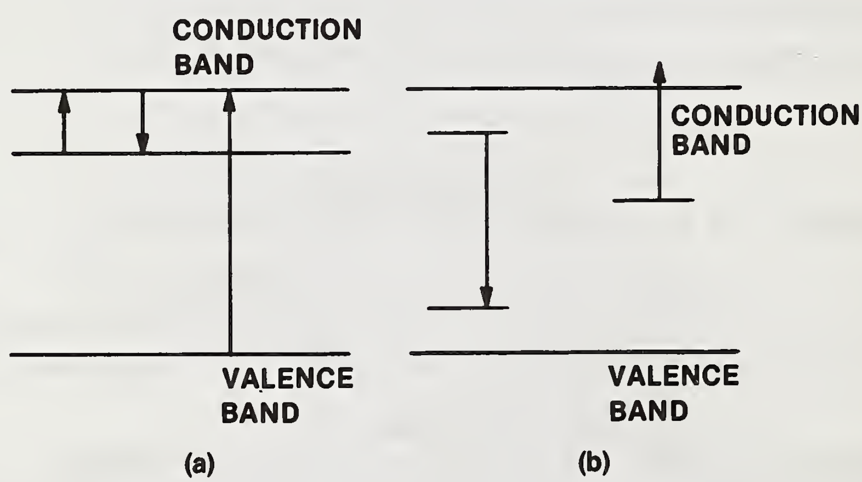

Figure 8. (a) Three-level semiconductor; (b) a phonon-assisted process.

---

[1]In some semiconductors; e.g., GaAs, band-to-band recombination can be important; in Ge and Si this process is improbable.

Figure 8(a), illustrating a three-level system, depicts an extrinsic semiconductor with electrons being generated from both the valence band and the impurity level and recombination taking place from the conduction band to the impurity level.

Figure 8(b), illustrating a more complicated many-level system, depicts a semiconductor with an electron making a transition to a lower energy level releasing phonon energy in the process which is absorbed by another electron, in another energy level, which then makes a transition up into the conduction band; i.e., the latter transition is a phonon-assisted transition.

The derivation of equations for fluctuations in a three-level system is already complex, and the complexity increases inordinately with each additional level assumed to be involved in the transition process.

Van Vliet and Fassett [19] have solved a three-level system problem where transitions are assumed to occur between the two bands and one set of impurity levels (or traps, or recombination centers). This can represent, for example, the case of indirect recombination in a semiconductor with generation taking place from the valence band to the conduction band, and recombination taking place via recombination centers. (In contrast, only direct recombination can occur in an assumed two-level system.) Using mathematical methods similar to those used for the two-level system, outlined above, for this particular case of indirect recombination, an expression of the following form is obtained for the noise spectra $G(\omega)$:

$$G(\omega) \simeq \frac{4\,\bar{n}\,\bar{p}}{\bar{n} + \bar{p}} \cdot \frac{\tau}{1 + \omega^2\tau^2} + \delta, \qquad (14)$$

where ω is (angular) frequency; $\bar{n}$ and $\bar{p}$ denote respectively the average number of electrons in the conduction band and holes in the valence band; τ is a composite relaxation time for the carriers; i.e., a complicated function of the numbers of each type of carrier and the relaxation times of each, and δ is a small term. If δ were neglected, the above equation would resemble that for g-r noise spectra in an intrinsic, two-level semiconductor.

4. Noise Power, Noise Temperature, and Noise Factor

In the measurement of small signals, electron devices are normally used in conjunction with amplifiers, and it is often desirable to know the noise limitation imposed by each. The noise of a device or component may be characterized by its noise power, noise temperature, or simply, noise, while it is convenient to characterize the noise of an amplifier (which depends on the signal) by its noise factor (a.k.a. noise figure). These specialized terms are defined forthwith.

4.1 Noise Power

Nyquist [11] showed that the available noise power P from a resistor R (into a "cold", or "noiseless", load) in an incremental band of frequency Δf is given by $P=kT\Delta f$.

Comparison of this expression with those for thermal noise [eqs. (12) and (13)] indicates that the available noise power may be given also by

$$P = \frac{\overline{v^2}}{4R}, \tag{15}$$

and

$$P = \frac{\overline{i^2}R}{4}. \tag{16}$$

The two expressions immediately above have been shown [21] to have general validity; i.e., they hold for any impedance, not just a pure resistance, and for any type of noise, not just thermal noise.

4.2 Noise Temperature

Noise temperature t_N is defined as the ratio of the available noise power output of a device or component to that of a resistor at room temperature; i.e.,

$$t_N = \frac{P}{kT_o\Delta f}. \tag{17}$$

It is convenient to choose as a standard temperature T_o = 290°K, in which case $kT = 4 \times 10^{-21}$ joule.

Noise temperature is actually a ratio, not a temperature. The rationale for the name is that the product $t_N T_o$ is the temperature a passive resistor would have to have in order to generate as much thermal noise as the noise output of the device or component in question.

4.3 Noise Factor

Noise factor, or figure, is a quantity which compares the noise from an actual amplifier with that from an ideal, i.e. noiseless, amplifier. This ideal amplifier is assumed to be identical to the actual amplifier, with the sole exception that it does not add any noise to the signal. Thus, if a source resistance is connected to the input terminals of a noiseless amplifier, the noise power output of the noiseless amplifier will be due only to the thermal noise in the source resistance.

The noise factor NF of an actual amplifier is defined as the ratio of the noise power output of the actual amplifier to the noise power output of the ideal (noiseless) amplifier.

Equivalently:

The noise factor NF of an amplifier is the ratio of the noise power at the output of the amplifier to the available thermal noise power of the source resistance; i.e.,

$$NF = \frac{P}{kT\Delta f}, \tag{18}$$

where T is the absolute temperature of the source resistance and P is the noise power measured at the output (corrected for gain) of the amplifier.

Noise factor depends on the value of source resistance, so any specification of NF must include the value of the source resistance used for the measurement. It is customary to express NF in decibels.

Methods of measuring NF of an amplifier, along with pitfalls in measurement, are given elsewhere [22].

5. Noise Theorems

To calculate the total noise or the total available noise power from a combination of noise sources, the following theorems are used:

Theorem 1. If a number of noise sources of differential (dynamic) resistance ρ_k are connected in parallel, the mean squares of the noise currents are additive; i.e.,

$$\overline{i^2} = \sum_k \overline{i_k^{\,2}} \qquad \text{(parallel)}, \tag{19}$$

and the total available noise power P is

$$P = \rho \sum_k \frac{P_k}{\rho_k} \qquad \text{(parallel).} \qquad (20)$$

$$\frac{1}{\rho} = \sum_k \frac{1}{\rho_k}$$

Theorem 2. If a number of noise sources are in series, the mean squares of the emf's are additive; i.e.,

$$\overline{v^2} = \sum_k \overline{v_k{}^2} \qquad \text{(series),} \qquad (21)$$

and the total available noise power is given by

$$P = \frac{1}{\rho} \sum_k P_k \, \rho_k \qquad \text{(series).} \qquad (22)$$

$$\rho = \sum_k \rho_k$$

6. Crystal Rectifiers

6.1 Operation

Owing to its historical importance, the early "cat whisker" type of crystal rectifier serves here as a first example for theoretical analysis of noise in a solid state device. This simple device consists essentially of a semiconductor; e.g., germanium or silicon, contacted by a sharp metal point, or whisker.

As for any two dissimilar materials when a metal and a semiconductor are brought into contact there will be an initial transfer of electrons from the material of lower work function (in this case, the semiconductor) to that of higher work function in order to establish equilibrium, as required by thermodynamics. A result of this process is the establishment of a potential barrier at the interface of the metal and semiconductor. This potential barrier is called the contact potential barrier and its height, the contact potential difference, will be given roughly by the difference in the two work functions. It is a characteristic of

this barrier that when a voltage is applied across it, the current passed in one direction will be larger than that passed in the other, a property regarded as rectification. Regardless of the polarity of the applied voltage with respect to metal and semiconductor, an electron current will flow from metal to semiconductor and another from semiconductor to metal, and the net current will be the difference in the two currents of opposite directions.

The width of the contact potential barrier is quite small; in fact, less than the mean free path of the electrons, so collisions (by the carriers) in the barrier layer can be neglected. Carriers having sufficient kinetic energy swiftly cross the barrier unimpeded and are immediately collected, so there is no interfering space-charge build-up.

The process wherein electrons traverse a potential barrier is termed emission, irrespective of whether the electrons emerge into a vacuum or another region of the solid state device. The latter process is called internal electron emission to distinguish it from the former. In the particular case of the crystal rectifier under discussion, the nature of the emission current is analagous to that of the temperature-limited vacuum diode, previously treated: the electrons are emitted randomly and remain free during transit. Hence just as for the vacuum diode, shot noise is expected to occur in the crystal rectifier. (In general shot noise may be associated with solid-state junctions; e.g., pn junctions and planar metal-semiconductor junctions.)

In addition to the rectifying barrier, there is a spreading resistance r in the semiconductor resulting from the constriction of current-flow lines in the semiconductor adjacent to the contact. When a voltage is applied across the device, there will be an Ir drop across this resistance, with the remainder of the applied voltage falling across the barrier.

6.2 Shot- and Thermal Noise

In the present analysis[1], for simplicity, only shot noise produced by the barrier and thermal noise caused by the ohmic spreading resistance, will be considered; further, the contact potential barrier will be assumed to be uniform, and the excitation will be assumed to be d.c. Figure 9 represents the equivalent circuit of the white noise components of the crystal rectifier.

[1]Based on the treatment of Torrey and Whitmer, loc. cit.

Figure 9. Schematic diagram of the equivalent circuit of a crystal rectifier; case of uniform contact potential.

The shot noise equation for the barrier is

$$\overline{i^2} = 2\ e\ I\ \Delta f, \tag{23}$$

where

$$I = I_1 + I_2, \tag{24}$$

I_1 and I_2 being respectively the electron current from the semiconductor to the metal, and from the metal to the semiconductor. The currents add because noise is produced each way.

The available noise power P_b from the barrier alone is, by eq. (16)

$$P_b = \tfrac{1}{2}\, eIR\Delta f, \tag{25}$$

where R is the differential resistance of the barrier.

The available noise power P_r from the spreading resistance r is

$$P_r = kT\Delta f. \tag{26}$$

The total available noise power P, obtained by using eq. (22), is

$$P = \frac{\tfrac{1}{2}\, eIR^2 + kTr}{R + r}\ \Delta f. \tag{27}$$

The noise temperature t_N is, by definition,

$$t_N = \frac{P}{kT_o\ \Delta f}, \tag{28}$$

where T_o = 290K. Therefore,

$$t_N = \frac{\frac{e}{2kT_o} \; IR^2 + r(\frac{T}{T_o})}{R + r} . \tag{29}$$

The net current I' traversing the barrier is given by the approximation [22]

$$I' \simeq A \, [e^{\alpha eV/kT} - 1] , \tag{30}$$

where A and α are constants incorporating the contact potential difference, and V is the voltage across the barrier.

On the other hand, I is given by

$$I \simeq A \, [e^{\alpha eV/kT} + 1] . \tag{31}$$

The differential resistance R of the barrier is

$$R = \left(\frac{dI}{dV}\right)^{-1} . \tag{32}$$

Therefore, from eqs. (30), (31), (32), and (25), the following formula is obtained for the available noise power from the barrier:

$$P_b = \tfrac{1}{2} \, kT\Delta f \;\; [1 + e^{-eV/kT}] . \tag{33}$$

Three cases may be demarcated:

1) $V = 0$; $\qquad P_b = kT\Delta f$ (34)

2) $V > 0$ and large; $\qquad P_b = \tfrac{1}{2} \, kT\Delta f$ (35)

3) $V < 0$ and large; $\qquad P_b \simeq \tfrac{1}{2} \, kT\Delta f[e^{-eV/kT}]$. (36)

For these formulas to hold, V must be less than the contact potential difference; otherwise the barrier no longer contributes noise.

6.3 Comparison of Theory and Experiment

The equations for t_N and P_b (noise temperature and noise power) are only in qualitative agreement with experiment, the measured noise being higher than that predicted by theory.

A simple model was used initially to avoid undue complexities in the mathematical derivations, but now the need for extending the model has been indicated. Previously, the effect of image force on barrier lowering was neglected; however, if this effect is included, the agreement is only slightly better. The assumption of a uniform contact potential was another simplification. Owing to unavoidable surface contamination and defects, the work functions and the contact area are not uniform. Hence, the equivalent circuit of the white noise components is more realistically represented by a distributed contact potential, as shown in figure 10.

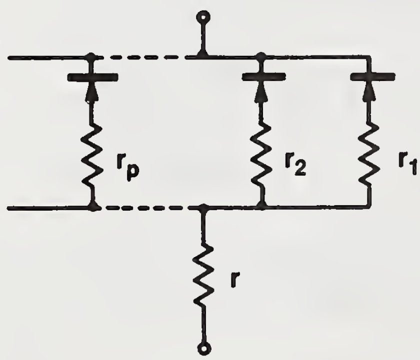

Figure 10. Schematic diagram of the equivalent circuit of a crystal rectifier; case of a distributed contact potential (Torrey and Whitmer [21]).

Now the total noise power P_b of the barrier must be found by using both noise theorems (sec. 5) to sum the contribution of the rectifying elements of different contact potential with their associated spreading resistance. Thus theory can be reconciled with experiment, but the elemental contacts are not unique, and their parameters difficult to verify. The solution to the distributed contact potential problem is complicated and its presentation here not justified.

Even the distributed contact potential model, already complicated, is a crude representation because the contact potential barrier has been tacitly assumed to remain invariant with time. Actually, there are electrons within the barrier layer constituting its space charge; since they are discrete particles they cause fluctuations in the height and width of the barrier. Current fluctuations, or noise, produced in a fluctuating barrier is expected to be enhanced compared to that produced in a static barrier, which perhaps can account for the discrepancy between existing theory and experiment.

7. Junction Transistors and Amplifiers

7.1 Operation of Junction Transistors

This section treats three aspects of a junction transistor: first, a rough description of the operation of a junction transistor is given; next, noise in a (two) junction transistor; finally, noise in a one-stage, low-frequency, low-noise amplifier made with this transistor.

The transistor, or solid-state triode, evolved from the point contact (whisker) crystal rectifier. In its initial form, the transistor consisted of two closely-spaced rectifying point contacts on a semiconductor with a large-area, non-rectifying contact across the base. In operation, one point emits charge (minority carriers) into the semiconductor crystal, while the other point collects the charge; the collector current I_c is controlled by the emitter current I_e, an effect known as transistor action. Figure 11 shows a set of typical characteristics of an n-type semiconductor point-contact transistor; V_c is the collector voltage.

A key parameter of the transistor is alpha (α), the grounded-base current gain, defined as $\alpha \equiv \partial I_C / \partial I_E$, at constant V_C. Transistors can be used to produce an amplifier exhibiting current gain, voltage gain, and power gain (as the dynamic resistance of the collector remains high even though the total current in the collector is markedly increased).

Rectifying contacts can be obtained also in a single chip of semiconductor crystal without the use of metal points. Here, the junction of two contiguous regions of opposite conductivity type (n and p) in a single crystal of semiconductor forms a rectifier with diode characteristics similar to that of the point-contact rectifiers. Thus, junction diodes can be made with a single pn junction, and junction transistors with a pair of back-to-back junctions (npn or pnp) on a single crystal--one conductivity type being sandwiched between the ends of the other conductivity type. Owing to their superior performance, stability, and quality control, junction devices have replaced most point-contact devices; nevertheless, point contacts remain viable, being found in the latest exotic devices.

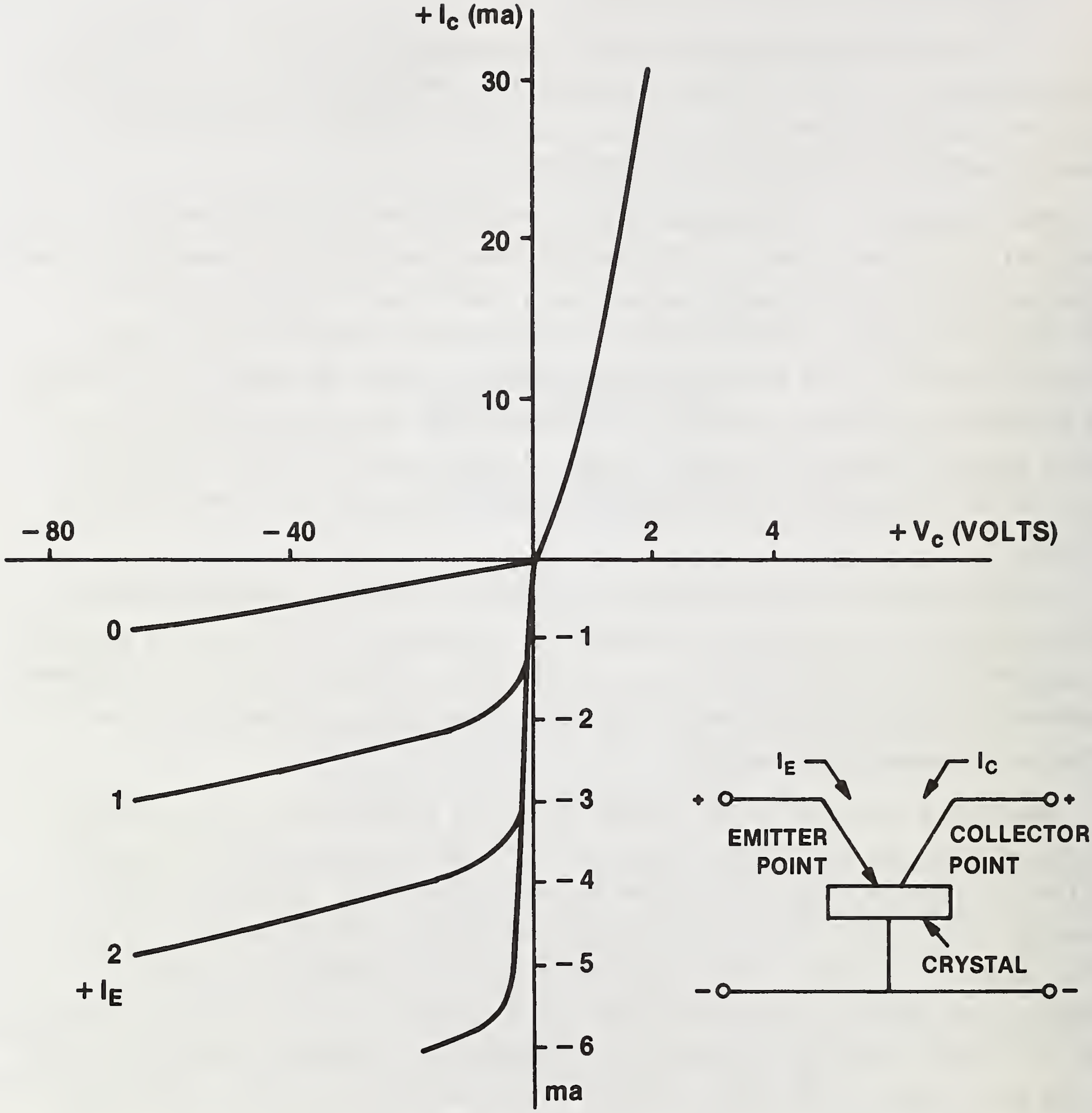

Figure 11. Characteristics of typical n-type semiconductor point-contact transistor (Hunter [23]).

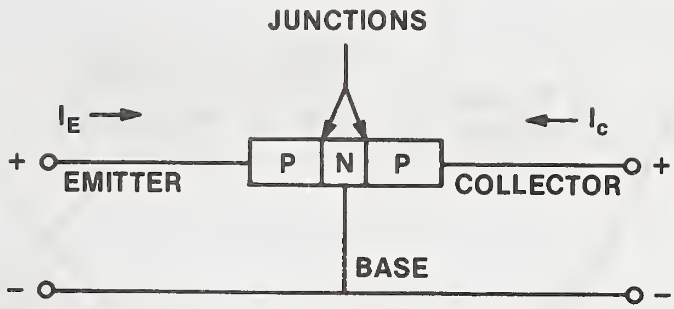

Figure 12. Schematic diagram of two-junction PNP transistor.

Figure 12 is a schematic diagram of a two-junction pnp transistor, and an ohmic contact is made to each region. Such a device may be viewed as two back-to-back diodes, or rectifiers, separated by a thin base region. The current-voltage characteristics of this transistor are similar to that of the point-contact transistor.

Junctions can be formed by any of several different methods; e.g., during growth of the semiconductor crystal, by alloying, or by diffusion of impurities into the crystal. The method of preparation affects the configuration of the junction--e.g., alloy junctions are abrupt and grown junctions tend to be diffuse--and thus affects the performance of the device, including the noise.

7.2 Noise in Junction Transistors

As for junction devices in general, three types of noise are manifest in junction transistors: 1/f noise, shot noise, and thermal noise. Shot noise and thermal noise are present at all frequencies, and 1/f noise is present at low frequencies; the type of noise which predominates depends on the frequency.

7.2.1 Frequency Spectrum

Figure 13 depicts the frequency spectrum of the noise factor NF, and three regions may be demarcated relative to frequency. In the low-frequency region 1/f noise predominates and the NF is seen to decrease linearly with increasing log frequency. At intermediate frequencies, the (frequency-independent) white noise, comprised of shot noise in the junctions and thermal noise in the base region, is predominant. At high frequencies, shot noise in the collector junction predominates. The white noise region is the region of minimum noise.

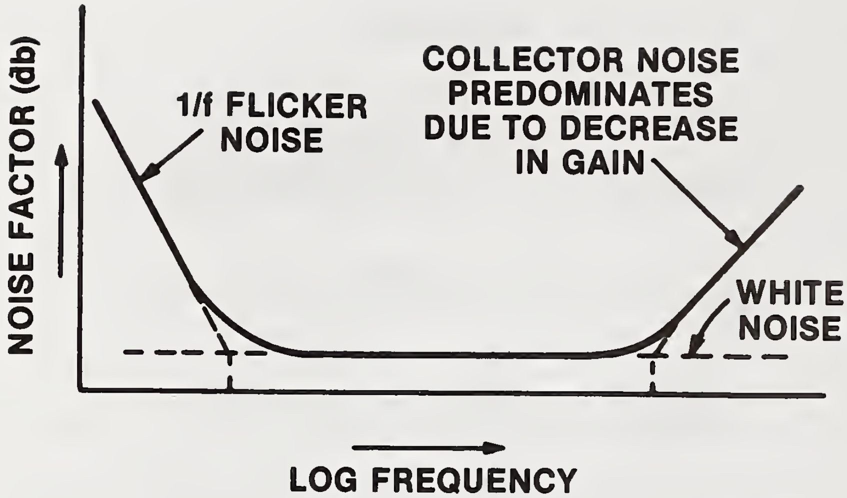

Figure 13. Noise factor vs. frequency (Woll and Herscher [22]).

7.2.2 Equivalent-circuit Analysis of White Noise

The equivalent circuit for the white-noise components of the junction transistor is represented by figure 14. The shot-noise components may be attributed to two correlated current generators, one at the emitter and one at the collector; however, these generators may be assumed to be independent, as it simplifies the mathematics and does not result in significant error when dealing with low-noise circuits. Thermal noise in the base is represented by an independent voltage generator in series with a noiseless base resistance $r_{bb'}$ (includes

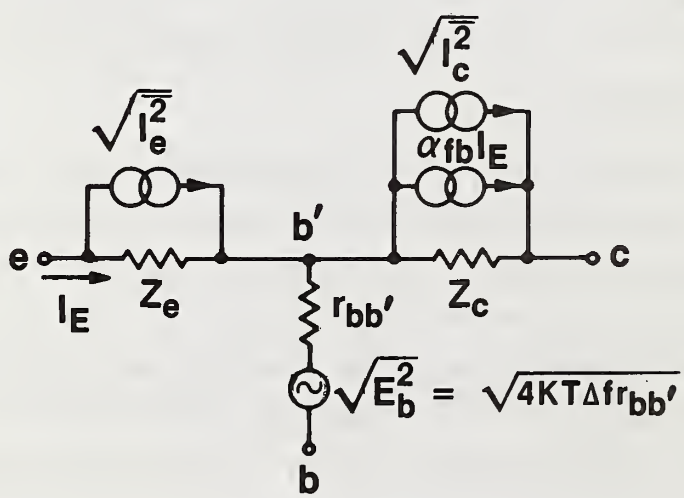

Figure 14. Equivalent circuit for white-noise sources (Woll and Herscher [22]).

the contact resistance to the base region as well as the innate resistance of the base region itself); cf figure 6.

The thermal-noise component is given by the familiar expression [eq. (12)]

$$\overline{v^2} = 4kTR\Delta f \tag{37}$$

where R is the differential resistance.

The shot-noise components are shown [24] to be

$$\overline{i_e^{\,2}} = 2eI_E\Delta f \tag{38}$$

and

$$\overline{i_c^{\,2}} = 2e\Delta f\,[I_E\,(\alpha_o - |\alpha_{fb}|^2) + I_{CO}] \tag{39}$$

$$= 2e\,\Delta f\,[I_C\,(1 - \frac{|\alpha_{fb}|^2}{\alpha_o}) + \frac{|\alpha_{fb}|^2}{\alpha_o}\,I_{CO}] \tag{40}$$

where i_e is the shot noise generated in the emitter,

i_c is the shot noise generated in the collector,

I_E is the direct emitter current,

I_C is the direct collector current,

I_{CO} is the reverse collector-to-base current with the emitter open-circuited,

α_o is the zero-frequency grounded-, or common-base current gain,

and α_{fb} is the short-circuit common-base current gain.

Also, because

$$\alpha_{fb} \simeq \frac{\alpha_o}{1 + (j\,f/f_{\alpha b})}\,, \tag{41}$$

where $f_{\alpha b}$ is the frequency at which $\alpha_{fb} = 0.707\alpha_o$, it can be shown from the above equations, that for f large, $\overline{i_c^{\,2}}$ increases with frequency. At low frequencies,

$$\alpha_{fb} \simeq \alpha_o\,, \tag{42}$$

and $$\overline{i_c^2} \simeq 2e\Delta f\, [I_c\, (1-\alpha_o) + I_{CO}\, \alpha_o] \, . \qquad (43)$$

Thus, for low noise, transistors having low 1/f cutoff frequency, high α_o, and high $f_{\alpha b}$ should be used ($I_{CO} << I_c$).

7.3 Transistor Amplifier

7.3.1 Types of Connection

A transistor can be used to build a single-stage amplifier, and the transistor can be connected in three different ways:

1) common-emitter, where the signal is applied to the base and taken from the collector;

2) common-base, where the signal is applied to the emitter and taken from the collector;

3) common-collector, or emitter follower, where the signal is applied to the base and taken from the emitter.

The type of connection made will determine the characteristics of the amplifier, and is chosen on the basis of intended application; e.g., the common-emitter connection gives the highest power gain of the three, making it generally the most suitable for use as an audio amplifier; the common-base connection has the lowest input resistance and the highest output resistance; the common-collector connection has approximately unity voltage gain, high input resistance, and low output resistance (making it analagous to a cathode follower).

An amplifier has an impedance associated with its input, called the input impedance, and an impedance associated with its output, called the output impedance. These two impedances are integrated within and internal to the amplifier. The external impedance which is connected to the input of the amplifier is called the source impedance and the external impedance which is connected to the output of the amplifier is called the load impedance.

Besides selecting low-noise transistors; i.e., those having the appropriate parameters, one wishes to know the optimum type of connection, operating point, and source impedance for minimal noise in a transistor amplifier.

7.3.2 Noise in Transistor Amplifiers

In general, below the alpha-cutoff frequency, the NF is found to be essentially independent of collector voltage and of type of connection; but is affected by the source impedance [22]. Provided sufficient gain can be obtained

in the first stage of a low-noise amplifier, the noise contributions from any succeeding stages can be neglected; because the source impedance yielding the minimum NF is nearly the same as that which provides maximum common-emitter gain, the common-emitter configuration usually is preferable.

As alluded to earlier, it is usually tacit in the term, noise factor, that the source impedance is a pure resistance. If the source impedance is not a pure resistance, the noise of the amplifier will be a function of frequency. If the source impedance is complex, the signal-to-noise ratio at the output is usually given, rather than the NF. If the source impedance is a pure reactance, the concept of NF is meaningless.

8. Solid-State Photon Detectors

8.1 General Types of Photodetector

There are two general types of solid-state photon detector; photoconductive and photovoltaic. In a photoconductive detector, light is used to produce a current in addition to that which would flow in the dark (i.e., with no illumination), and the device must be biased in order to operate. In a photovoltaic detector, light is used to produce a voltage, and no bias, or applied voltage, is required. The properties of these two types of detector differ, and the choice is governed by intended application; e.g., photovoltaic cells tend to have larger active areas and may be used directly to power devices.

Both photoconductive- and photovoltaic detectors generally utilize semiconductors (sometimes insulators are used), but differ in configuration. Examples of several photoconductive- and photovoltaic configurations are shown in figure 15. A photoconductive detector may be a single crystal slab of semiconductor (or insulator) with electrical contacts at two opposite faces, such as depicted in (a), or a polycrystalline thin film of semiconductor deposited on an insulating substrate (e.g., by evaporation, sputtering, or electrolysis), as depicted in (b). Photovoltaic detectors are either metal-semiconductor barrier devices (c) or junction devices (d) of planar geometry. For the barrier devices, a "semitransparent" metal film, usually of gold, deposited on the surface of the semiconductor crystal, makes electrical contact to the crystal while transmitting roughly half the radiation which impinges on it. The interface of the gold film and the semiconductor forms a rectifying barrier (somewhat similar to that formed by a metal point contact). Such a device is called a Schottky barrier photodiode[1].

[1]Some authors reserve the term photodiode to signify an externally-biased photovoltaic detector.

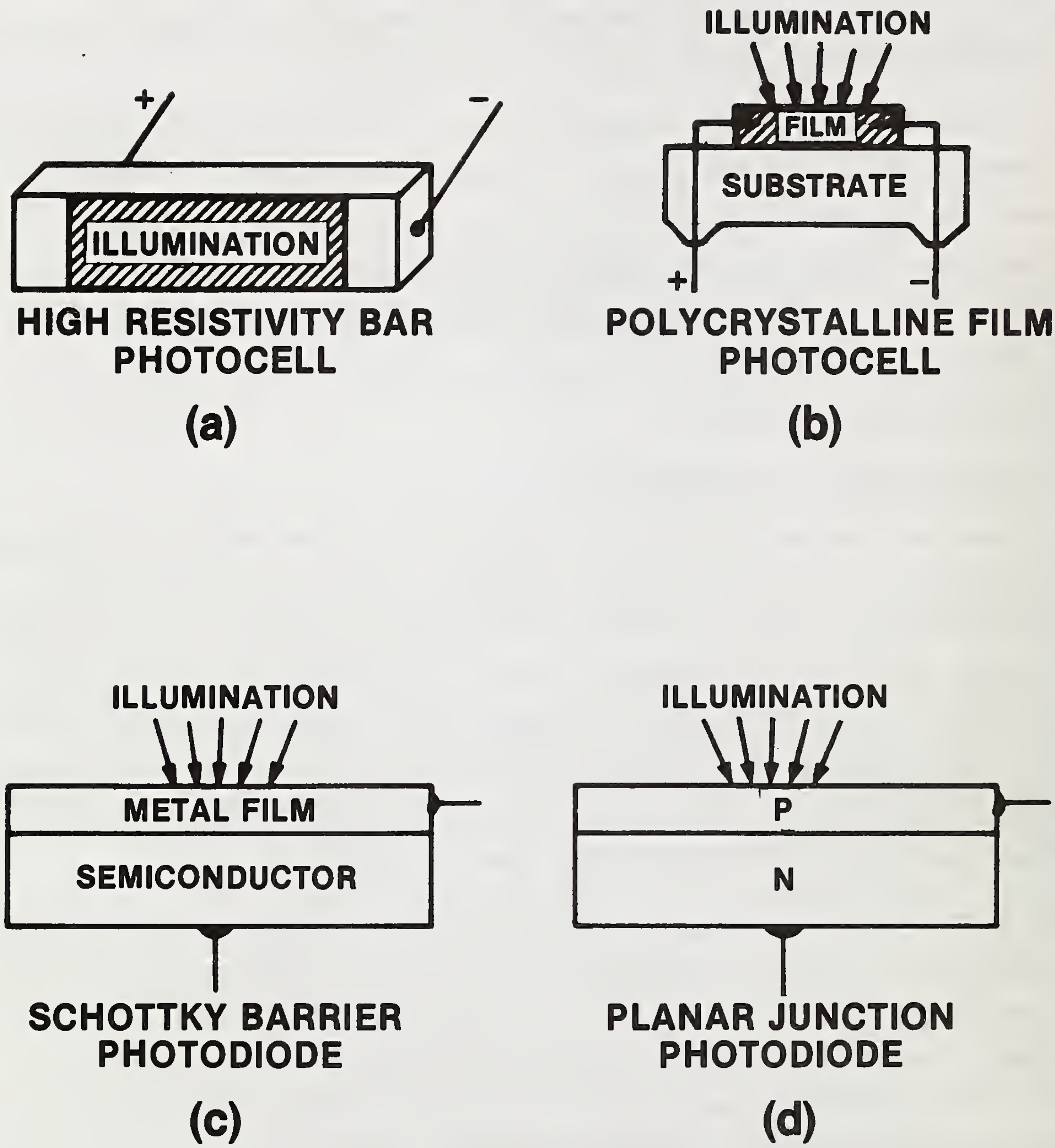

Figure 15. Various types of photocells.

Alternatively, a (rectifying) planar pn junction near the surface of a single crystal of semiconductor, as shown in (d), may be fabricated. Such a device is called a planar junction photodiode.

8.2 The Photoconductive Process .

The photoconductive process was alluded to earlier in the treatment of g-r noise, as both processes originate in the transitions of electrons between allowed energy levels. The essential differences in the two phenomena are that photoconductivity relates to the average rates of generation and recombination of the carriers and the generation transitions are signal photon induced; whereas g-r noise relates to fluctuations, or deviations, from the average generation and recombination rates, and the generation transitions may be either photon induced or thermally induced.

Photoconductivity may be more readily understood with the aid of figure 16, a schematic representation of the pertinent energy levels in the two broad classes of semiconductor, intrinsic and extrinsic. In intrinsic semiconductors the concentration of impurities which can act as either donors or acceptors of electrons is negligible; in extrinsic semiconductors this impurity concentration is preponderant. In figure 16, (a) relates to an intrinsic semiconductor, (b) to an n-type, or donor, extrinsic semiconductor, and (c) to a p-type, or acceptor, extrinsic semiconductor.

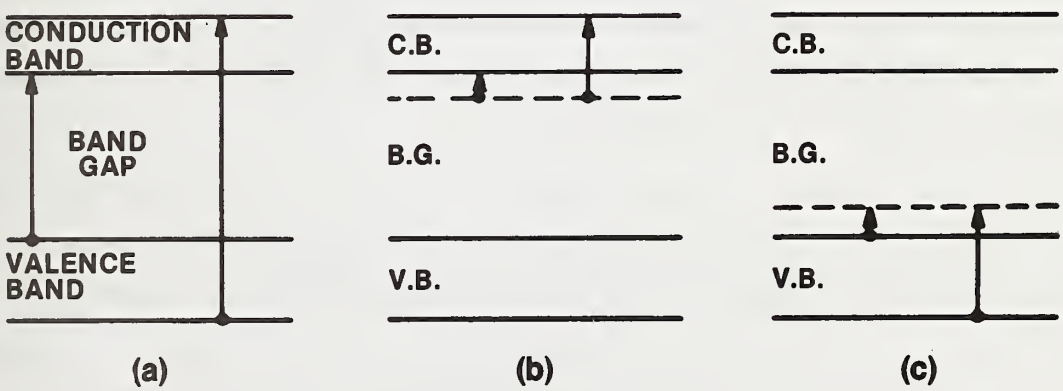

Figure 16. Energy band diagrams illustrating photoexcitation in intrinsic (a) and extrinsic (b) and (c) semiconductors.

A semiconductor at, say, room temperature and in the dark already has carriers--free electrons and holes--owing to thermal excitation of electrons and background radiation. In the intrinsic semiconductor electrons are thermally excited by the lattice vibrations, from mainly the top of the valence band to the bottom of the conduction band. The electrons in the conduction band are mobile, as are the electron vacancies--the holes--left behind. Thus, when a voltage is applied across the semiconductor a (dark) current flows. The situation for an extrinsic semiconductor, figure 16 (b) and (c), is somewhat similar, except that the electrons make transitions between a band and impurity levels, rather than between bands and only one type of carrier (the majority carrier) need be considered. In either case recombination of carriers ensues, the rate of generation being equal to the rate of recombination as required for equilibrium.

Shining light of sufficient energy on the semiconductor will excite additional electrons into making transitions, either interband or between an impurity level and a band, thus creating additional (mobile) charge carriers--hole-electron pairs in intrinsic material and mobile holes <u>or</u> electrons in extrinsic material--which in turn increases the current flowing in the semiconductor. Recombination of these additional carriers also ensues. The vertical arrows in figure 16 indicate the direction and location of the photon-induced electron transitions. The energy of a photon is $h\nu$, where h is Planck's constant and ν is the frequency of the light. Thus, if E_i denotes the difference in energy between the two levels involved in a transition, the frequency of the light required to effect the transition when absorbed in the material will be given by $\nu_i = E_i/h$. In the figure, the short arrows relate to the effective minimum frequencies (or maximum wavelengths) and the long arrows to the effective maximum frequencies (or minimum wavelengths) required to obtain photoconductivity.

As for the dark case, the rates of electron generation and hole generation produced by the light are equal for intrinsic material, and the rate of generation of a hole or electron is equal to the rate of recombination of a respective hole or electron.

After the light has been turned on sufficiently long for the equilibrium photocurrent to be established and the light is turned off "instantaneously", there will be an exponential decay in the photocurrent. The time after turn-off for the photocurrent to decay to 1/e (37 percent) of its steady-state value is the lifetime, and the corresponding distance traversed by the carriers is the diffusion length. The values of these parameters do not depend on turning off the

light. They remain the same when the light is on; turning off the light is used here merely for illustrative purposes.

8.3 Measurements of Photoconductivity

Photoconductive detectors of visible radiation ("visible detectors") have relatively large band gaps, or activation energies, so the number of carriers in the dark tends to be small and the number of photon-excited carriers large--hence photoconductivity may be observed with steady irradiation and with the detector at room temperature.

Infrared photoconductors tend to have small band gaps (if intrinsic) or small impurity activation energies (if extrinsic), and the number of carriers in the dark tends to be large, with correspondingly large dark current. On the other hand, the number of photon-excited carriers is small, and their presence probably would not be detectable if special measures were not taken. These include cooling the detector in order to reduce the dark current and the noise, and/or modulating the radiation in order to produce a modulated photocurrent which then may be measured by an amplifier sharply tuned to the frequency at which the radiation is being modulated.

An important performance parameter of a photodetector is its responsivity, which expresses its output signal current or voltage per watt of incident radiant power. For detectors of visible radiation this parameter is usually adequate because the signal current tends to be large compared to the dark current, and noise is negligible. At low levels of radiant power, however, the noise becomes appreciable relative to the small-signal output from the detector and signal-to-noise ratio is an additional, and sometimes more useful, performance parameter.

8.4 Infrared Detectors

8.4.1 Detectivity

In infrared detectors, where the signal outputs are small, noise is always an important consideration, and the ratio of responsivity to noise defines the detectivity of the detector. For convenience, detectivity is usually normalized to a detector area of 1 cm^2 and an amplifier bandwidth of 1 Hz; the normalized detectivity is written D* (pronounced Dee-star). Detectivity measurements of photoconductive infrared detectors are made by connecting the detector in series with a load resistor and a battery (photovoltaic detectors may be measured by connecting them directly to the preamplifier); a blackbody simulator, usually at 500 K, provides the signal radiation, which is then modulated sinusoidally with a mechanical chopper assembly, usually at 900 Hz. Thus D* may be written

$$D^*(500K,\ 900\ Hz) = \frac{\frac{S}{N}(\Delta f)^{\frac{1}{2}}}{H(A)^{\frac{1}{2}}}\ , \quad (44)$$

where S is the signal measured across the load resistor with signal radiation on the detector, N is the noise measured across the load resistor in the dark (i.e., without signal radiation on the detector), Δf is the band pass of the amplifier, H is the power density of the signal radiation in Wcm^{-2}, and A is the area of the detector in cm^2.

8.4.2 Small-signal Circuit Analysis

The equivalent small-signal circuit of the detector may be given by a current generator shunted by both the detector resistance R_D and the load resistor R_L, where the detector has been represented by a short-circuit current generator in shunt with the detector resistance. Then, the short-circuit current responsivity $\mathcal{R}_{sci}$ can be calculated from the following relation [25]:

$$\mathcal{R}_{sci} = \frac{\eta e \lambda G_p}{hc}(1 + \omega^2\tau^2)^{-\frac{1}{2}} \quad (45)$$

where η is the quantum efficiency, e the charge of the electron, λ the wavelength of irradiation, h Planck's constant, c the velocity of light, G_p the photoconductive gain, $\omega = 2\pi f$ where f is the modulation frequency, and τ the carrier lifetime. At low frequencies or dc where $\omega^2\tau^2 << 1$.

$$\mathcal{R}_{sci} \simeq \frac{\eta e \lambda G_p}{hc}\ . \quad (46)$$

The photoconductive gain G_p is given by

$$G_p = \tau/t = \mu\tau E/l\ . \quad (47)$$

where τ is the carrier lifetime, t is the carrier transit time, μ is the carrier mobility, E is the electric field strength, and l is the interelectrode spacing.

The measured signal voltage across the load resistor S_v is

$$S_v = P\,\mathcal{R}_{sci}\left(\frac{R_D\ R_L}{R_D + R_L}\right), \quad (48)$$

where P is the radiant power on the detector in watts.

This circuit analysis is applicable to a photovoltaic detector as well as a photoconductive detector, except that the gain G_p of a photovoltaic detector is unity.

8.4.3 The Photovoltaic Effect

The electronic processes which occur in a solid when light interacts with a semiconductor are determined by the characteristics of the light (energy and spectral distribution) and the material parameters (e.g., light absorption coefficient, band gap energy, impurity concentration). Thus the physics of photoconductive- and photovoltaic detectors have much in common. Nevertheless, there are some notable differences in the two types of device which arise from the difference in their compositions: photoconductive detectors are homogeneous and may use intrinsic or extrinsic semiconductors; photovoltaic detectors are heterogeneous, incorporating a (rectifying) contact potential barrier in the form of a metal-semiconductor barrier or a pn junction; only intrinsic semiconductors may be used for Schottky photodiodes. Owing to the presence of this barrier/junction, no external bias is required for operation (as is the case for photoconductive detectors). Assuming zero bias for the photovoltaic detector, there will still be a thermal noise. An illuminated barrier/junction with zero bias applied may be likened to an unilluminated barrier with bias applied (see sec. 6); i.e., illuminating the barrier/junction is tantamount to applying an external voltage across the barrier/junction. The barrier/junction serves as a charge separator. For example, in the junction device, when hole-electron pairs are created through the absorption of light energy, the electrons at the junction will pass easily only in one direction and be collected virtually without recombination having occurred, while the holes will pass easily only in the opposite direction and be collected also without recombination having occurred. Thus it is important that the junction thickness be small compared to the mean free path of the carriers to prevent current-retarding space-charge buildup in the junction and that the light energy be absorbed within a diffusion length of the junction. Details of the current flow in a Schottky barrier photodiode differ[1], but the same restrictions apply to barrier width.

8.4.4 Noise in Photoconductive Detectors

Photoconductive detectors exhibit three of the four types of noise which were identified as principal noises: 1/f noise, thermal noise, and g-r

---

[1]See ref. 23, section 5.

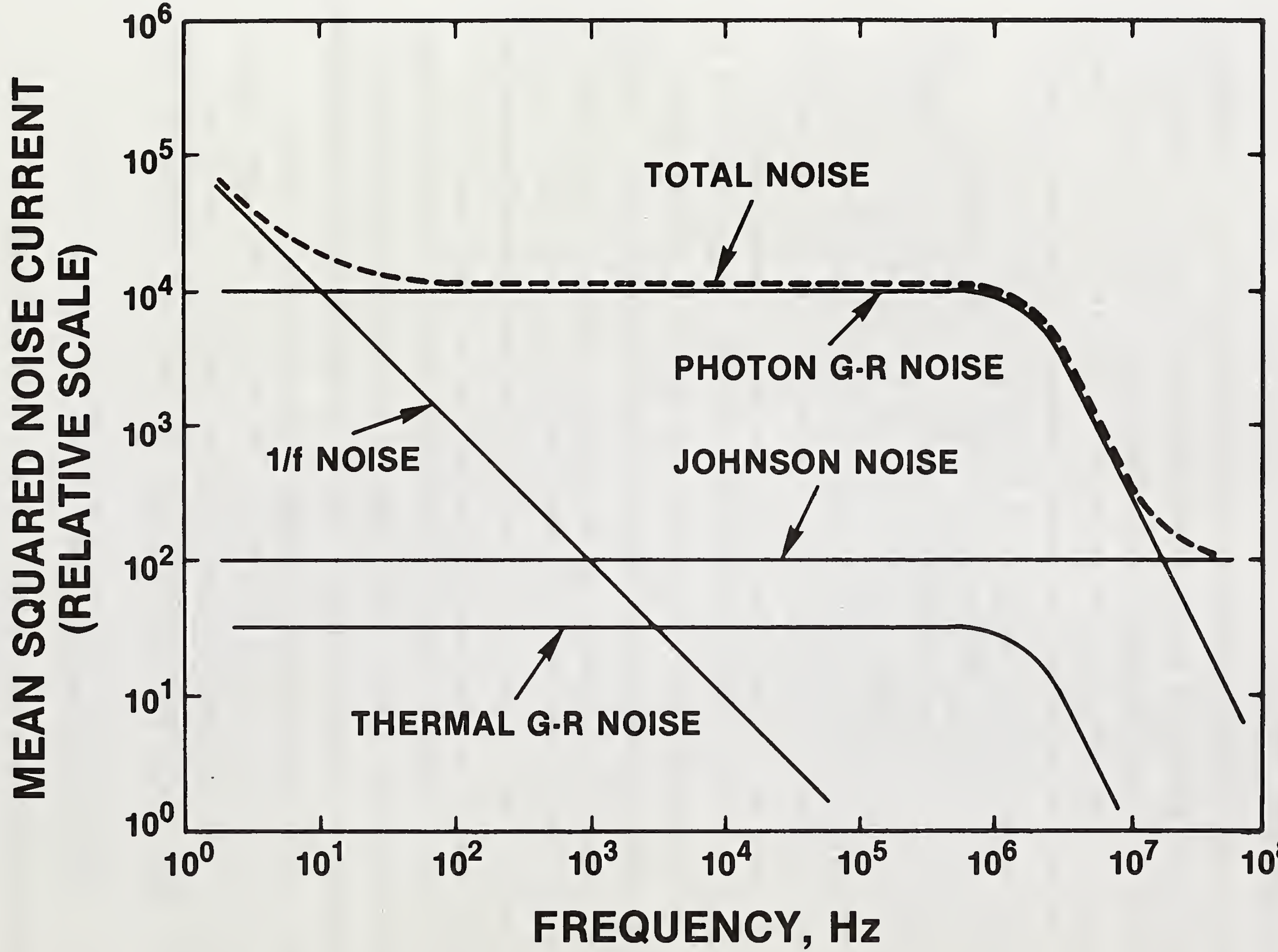

Figure 17. Illustration of the contribution of various noise sources to the total noise spectrum (Bratt [26]).

noise. Figure 17, a plot of the mean-squared noise current vs. (amplifier) frequency, illustrates how the various noises may be manifest[1].

Shown individually are the 1/f noise, prominent at low frequencies, but decreasing with increasing frequency[2]; the white Johnson, or thermal noise, present at all frequencies, but overshadowed by other noises; and the two forms of g-r noise--that which is produced by background radiation (photons) is commonly referred to as (background) photon noise to distinguish it from that produced by thermal vibrations of the lattice, referred to simply as g-r noise. An equivalent circuit would have each noise current source mutually parallel; thus the total noise spectrum is formed by addition of the individual noise currents. Three regions may be demarcated according to frequency range: At low frequencies 1/f noise predominates; at intermediate frequencies the noise is essentially photon noise and this is the frequency region of normal operation; at high frequencies the photon noise decreases and approaches thermal noise.

That the g-r noises are flat may be understood from the following: the spectral density of the noise current $S_i(f)$ is defined by the relationship

$$\overline{i^2} = \int_0^\infty S_i(f)\, df \;, \tag{49}$$

where $\overline{i^2}$ is the mean squared short-circuit noise current and f is frequency. (The frequency dependence of the noise current is implicit in $S_i(f)$; e.g., for thermal noise, $S_i(f) = 4kTG$.)

For the general case of g-r noise (thermal) in an extrinsic semiconductor containing both donor impurity atoms and acceptor impurity atoms, but with one in the majority, $S_i(f)$ is a function of the total number of free electrons N, the total number of acceptor impurity atoms N_a, and the total number of donor impurity atoms N_d [19,27]. Extrinsic photoconductive detectors are normally operated cooled, the result being that $N << N_a, N_d$. Then $S_i(f)$ is given by

---

[1]This example represents cooled, extrinsic silicon and germanium detectors; for other detectors or other operating conditions, the relative magnitudes of the noise sources may differ.

[2]1/f noise in germanium devices is attributed to an inversion layer at the surface, while in silicon devices it is associated with the oxide passivating layer.

$$S_i(f) = 4\,\frac{I^2}{N}\left[\frac{\tau}{1 + (2\pi f\tau)^2}\right] , \tag{50}$$

where I is the steady, or average, current and τ is the lifetime of the free carriers. Thus at low frequencies where $2\pi f\tau << 1$, the noise is flat, or frequency independent; when $2\pi f\tau >> 1$, the noise decreases. At high temperatures where $N \approx N_d - N_a$ (N_d has been assumed to be the major impurity), the g-r noise associated with impurities approaches zero, but g-r noise from the now intrinsic semiconductor becomes significant and will escalate with increasing temperature.

In most cases photon noise will be equivalent to (thermal) g-r noise, and eq. (50) derived for the latter is applicable too for the former when N is replaced by N_b, the total number of free electrons generated by the background radiation. In those rare cases where detectors are operated at room temperature and used for long wavelengths $\gtrsim$ 20 µm, however, eq. (50) would have to be modified [26].

8.4.5 Noise in Photovoltaic Detectors

Photovoltaic detectors may exhibit all four principal noises, the additional noise being shot noise, if the detector is biased. The shot noise arises because of the contact potential barrier associated with either the Schottky metal-semiconductor barrier or the pn junction of the photovoltaic detector. If the contact potential barrier width is small--i.e., less than the mean free path of the conducting electrons--virtually all of these electrons will traverse the barrier without suffering collisions within the barrier and there will be no resultant space charge buildup. The current through the barrier will be the saturated current I_s, and full shot noise will be generated in the barrier. The equation for shot noise, repeated here for convenience, is

$$\overline{i^2} = 2eI_s\Delta f , \tag{51}$$

and the equivalent circuit, a current generator (shunted by an infinite impedance), has already been shown in figure 3 in connection with the thermionic diode. On the other hand, if the barrier width is not less than the mean free path, there will be a space-charge buildup in the barrier and the current through the barrier will be space-charge limited and reduced. As for the analogous thermionic diode, previously treated, the noise will be a partial shot noise substantially lower in magnitude than the full shot noise as would be given by the equation. And as for the thermonic diode, the magnitude of partial shot noise is unpredictable; the equivalent circuits of figure 4 are applicable here too.

There is a further important distinction between the photon noise in a photoconductive detector and in a biased photodiode. In a properly designed and constructed photodiode as described in the preceding section--i.e., one where the radiation is absorbed in a distance from the barrier/junction which is much less than the diffusion length, and the thickness of the barrier is less than the mean free path--there will be negligible recombination taking place[1]. Hence there is no recombination noise, just generation noise. In a photoconductive detector, however, the rates of generation and recombination are equal, both contributing equally to the noise. Thus the generation noise (g noise) power in a biased photodiode will be one-half the g-r noise power in the corresponding photoconductive detector, and the detectivity for a given incident energy flux for a photovoltaic background (photon) limited detector is therefore greater by a factor of 2 than for a corresponding photoconductive detector prepared from the same material. Otherwise, the noise spectrum of a cooled photodiode, particularly one of silicon or germanium, would be somewhat similar to that shown in figure 17 for a photoconductive detector.

8.4.6 The Ideal Detector

The higher the detectivity of an IR detector, the smaller the amount of radiant energy it is capable of detecting, and for a detector to realize its optimal detectivity, the limiting noise is required to be background photon noise; i.e., all other noises must be suppressed (e.g., by cooling the detector, modulating the radiation at high enough frequency to squelch 1/f noise, and using a large load resistance). Even so, no actual detector is perfect so the maximum detectivity achievable is always less than the maximum value of detectivity theoretically possible with an assumed perfect, or ideal, detector.

The fictitious ideal detector is background photon noise limited and has a quantum efficiency of unity up to the long wavelength cutoff and zero beyond (the latter stipulation is unrealistic). Detectivities may be calculated as function of wavelength for ideal detectors, and curves generated therefrom, an example of which is shown in figure 18. Such curves are useful for determining the highest value of detectivity theoretically possible at a given wavelength. Thus for example, an actual detector may approach, but not attain, a $D^*(\lambda)$ of 1×10^{11} cm $Hz^{\frac{1}{2}}$ W^{-1} at a wavelength $\lambda = 5$ μm, with conditions as specified in the figure.

---

[1]At zero-bias, however, there is no net current so the rates of recombination and generation must be equal.

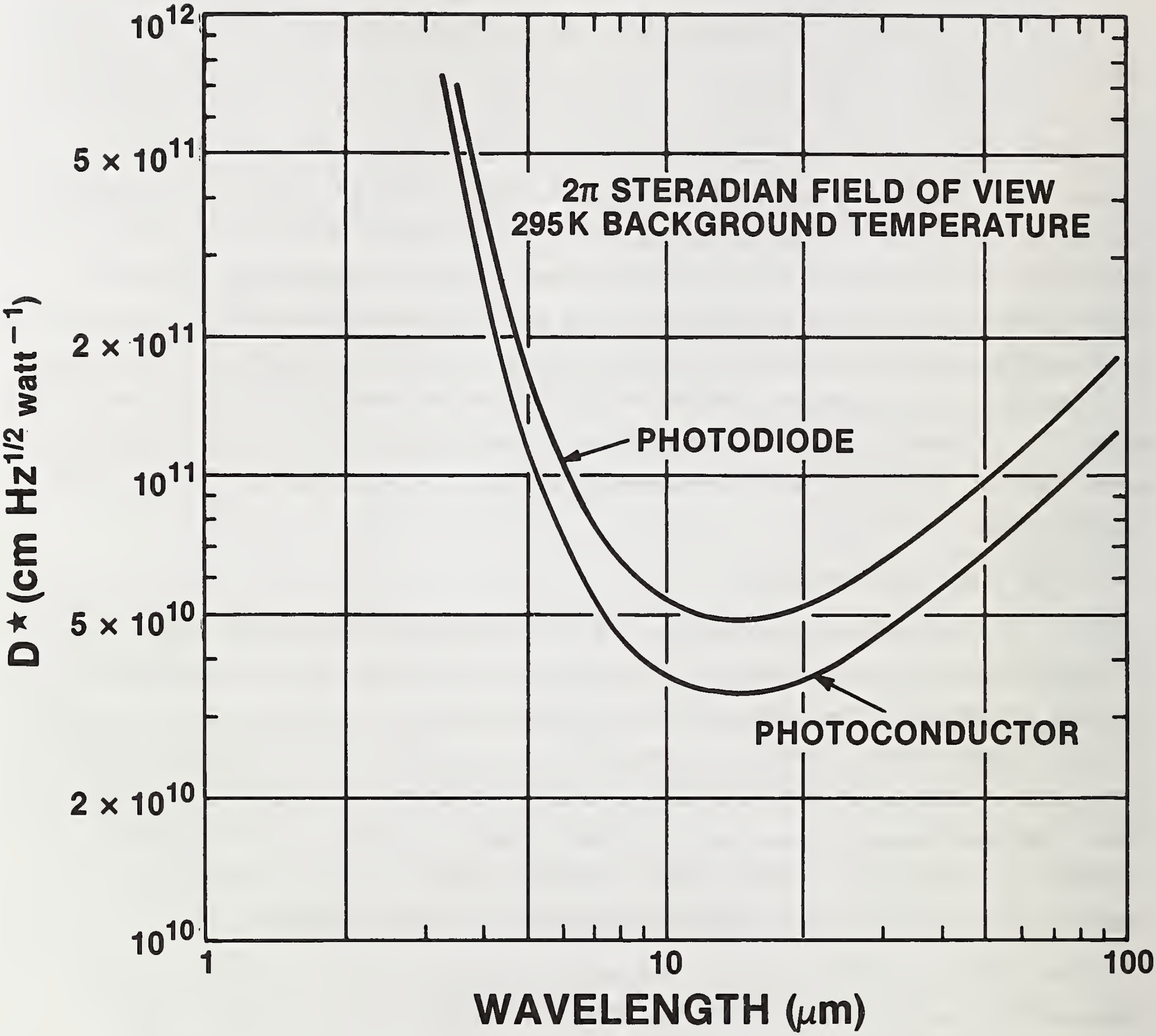

Figure 18. The dependence of the detectivity of an ideal background-limited detector on cutoff wavelength (Bode [25]).

Background-limited infrared photodetectors are said to be in the blip condition, blip being the acronym. Photon noise may be reduced (thus increasing D*) by reducing the field of view of the detector and by reducing the temperature of the background; e.g., by use of a cooled apertured baffle. Other ideal detector detectivity vs. wavelength curves covering various situations may be found elsewhere [26,28,29].

8.4.7 Amplifier-limited Noise

If the background noise of a detector in the blip condition is sufficiently reduced, the detectivity will become limited by noise in the preamplifier. This condition is referred to in the jargon as plip, the acronym for preamplifier limited infrared photodetector. If the amplifier noise is caused mainly by thermal noise in the source resistance, <u>increasing</u> its value will reduce the noise. This follows because in an equivalent circuit this noise source is represented by a current generator and noiseless source resistance in shunt with the amplifier input; the thermal noise current in the source resistance is given by eq. (13). See also figure 6 and section 3.2.7.

9. Concluding Remarks

Most of the ongoing work involving noise is aimed at its reduction or minimization, and either of two approaches may be taken: 1) device design, including new and improved materials; (2) optimal operation of given devices. The former, however, is an esoteric pursuit, and the latter accounts by far for most of the activity.

Throughout this work various operational means for reducing or minimizing noise have been given explicitly, while more may be easily inferred. To recapitulate: Obviously 1/f noise may be reduced by operating at higher frequencies, and by reducing current; shot noise in junction or barrier devices in general may be reduced by reducing current, and avoided or minimized in photodiodes by zero-bias operation; thermal noise may be reduced by reducing temperature and by increasing resistance; thermal g-r noise may reduced by reducing temperature; photon noise in IR detectors may be reduced by reducing field of view and background temperature; noise in general may be reduced by reducing bandwidth. Further, section 7.7.2 discusses requirements for low-noise junction transistors and amplifiers.

Sometimes it may not be feasible or desirable to employ one or another of these methods because of the consequences. For example, reducing bandwidth will reduce noise, but also may block out desired sidebands; or a device may be operated advantageously in a mode which increases noise, but which also increases the

signal-to-noise ratio, as when a photodiode is reverse-biased. In other words, in practical situations it is necessary to consider the trade-offs, if any.

A word about noise identification: The principal noises have distinguishing characteristics which may be used for such purpose.

Recapitulating: Thermal noise is the only noise which occurs in a passive resistance and which is (explicitly) independent of current. All other noises require the existence of steady current and depend on the current noise, increasing with increasing current. Of these other noises, only shot noise is proportional to the square root of the current, the remainder being proportional to the current. Thermal noise and shot noise are white noises. On the other hand, the distinguishing characteristic of 1/f noise is its frequency dependence ($\overline{i^2} \alpha$ 1/f). Photon noise is sensitive to background temperature and field of view. Identification of particular noises will be facilitated if the particular noise has predominance; otherwise it may be inextricable.

Finally, for those readers going on to investigate noise in electron/solid state devices: van der Ziel's recent historical survey of noise [30] provides an extensive source of references; the (out-of-print) book by Kruse et al [29] contains an anomalously lucid treatment of noise; and, of course, a detailed knowledge of the operation of the device of concern is requisite.

10. References

[1] Proc. 6th International Conference on Noise in Physical Systems, NBS Special Publications 614, Wash. D.C., (1981).

[2] G. Herrmann and S. Wagener, The Oxide Coated Cathode, Vol. 2, pp 10, 11, Chapman & Hall Ltd., London (1951).

[3] O. W. Richardson, Emission of Electricity from Hot Bodies, 2nd edition, London (1921).

[4] G. Herrmann and S. Wagener, The Oxide-Coated Cathode, Vol. 2, p 17, Chapman & Hall, London (1951).

[5] The Scientific Papers of J. Clerk Maxwell, W. D. Niven, ed., Vol. I, p 380, Cambridge Univ. Press (1890).

[6] W. Schottky, Ann. Phys. 57, 541 (1918).

[7] K. R. Spangenberg, Vacuum Tubes, p 306, McGraw-Hill Book Co., Inc. (1948).

[8] R. Brown, Phil. Mag. 4, 161 (1828).

[9] A. Einstein, Ann. Phys. 17, 549 (1905); Ann. Phys. 19, 289, 371 (1906).

[10] J. B. Johnson, Nature 119, 50 (Jan. 8, 1927); Phys. Rev. 29, 367 (Feb., 1927); Phys. Rev. 32, 97 (July 1928).

[11] H. Nyquist, Phys. Rev. 32, 110 (July 1928).

[12] J. Bernamont, Ann. Physique 7, 71 (1937).

[13] C. J. Bakker and G. Heller, Physica VI, no. 3, 262 (March 1939).

[14] D. A. Bell, J. Instn. Elect. Eng. 82, 522 (1938).

[15] D. A. Bell, loc. cit.; see also its appended Discussion.

[16] J. B. Johnson, Phys. Rev. 26, 71 (1925).

[17] H. N. Kozanowski and N. H. Williams, Phys. Rev. 36, 1314 (1930).

[18] W. Graffunder, Telefunken-Röhre 15, 41 (1939).

[19] K. M. van Vliet and J. R. Fassett, in Fluctuation Phenomena in Solids, R. E. Burgess, ed., Academic Press, New York, Chap. 7 (1965).

[20] References (59) through (66) in K. M. van Vliet and J. R. Fassett, loc. cit.

[21] H. C. Torrey and C. A. Whitmer, Crystal Rectifiers, p 26, McGraw-Hill Book Co., Inc. (1948).

[22] H. J. Woll and M. B. Herscher in Handbook of Semiconductor Electronics, L. P. Hunter, ed., 2d ed., pp 11-83, McGraw-Hill Book Co., Inc., New York (1962).

[23] L. P. Hunter in Handbook of Semiconductor Electronics, loc. cit., Section 1.

[24] A. van der Ziel, Proc. IRE 46, no. 6, pp 1019-1038 (June, 1958).

[25] D. E. Bode, "Infrared Detectors" in Applied Optics and Optical Engineering, Vol. VI, R. Kingslake and B. J. Thompson, eds., Chap. 8, Academic Press, New York (1980).

[26] P. R. Bratt, "Impurity Germanium and Silicon Detectors", in Semiconductors and Semimetals, Vol. 12, P. K. Willardson and A. C. Beer, eds., Chap. 2, Academic Press, New York (1977).

[27] K. M. van Vliet, Proc. IRE 46, 1005 (1958).

[28] K. M. van Vliet, Appl. Optics 6, 1145 (1967).

[29] P. W. Kruse, L. D. McGlauchlin, and R. B. McQuistan, Elements of Infrared Technology, Chapters 7 and 9, Wiley, New York (1962).

[30] A. van der Ziel, "History of Noise Research", in Advances in Electronics and Electron Physics, Vol. 50, L. Marton and C. Marton, eds., Academic Press, New York (1980).

[31] Cf S. O. Rice, Bell Syst. Tech. J. 23, 282 (1944).

[32] Cf R. V. Churchill, Fourier Series and Boundary Value Problems, p 43, McGraw-Hill Book Co., Inc., New York (1941).

APPENDIX

This appendix contains derivations of the thermal noise equation and the shot noise equation. Each is derivable by more than one way; the treatments here are based on Nyquist's [11] deduction for the thermal noise and on Schottky's [6] method for the shot noise. The former is mainly physical--although Nyquist assumed familiarity with properties of transmission lines--and gives insight into the nature of noise. On the other hand, the latter is mainly of mathematical interest.

A. Derivation of Thermal Noise

Nyquist's calculation of the electromotive force (emf) owing to thermal noise in a conductor cleverly couples principles in thermodynamics and statistical mechanics with principles in electricity, specifically those governing electrical transmission lines. In the present treatment properties of lines are first reviewed, after which the thermal noise emf is derived.

A line is a pair of long parallel conductors, each of uniform cross section, used to carry ac. A line has a resistance R_L and an inductance L per unit length (both sides of the line are included), and a leakage conductance G and a capacitance C between conductors per unit length. Thus the series impedance per unit length along the line is given by $Z_L = R_L + L\omega$ and the shunt impedance per unit length across the line by $Z_c = 1/(G + iC\omega)$. The characteristic impedance Z_K of a line refers to the impedance of the line if it is infinitely long. An infinite line can contain no reflected waves because the original transmitted wave is attenuated to zero before it reaches the output end, and

$$Z_K = (Z_L\ Z_C)^{\frac{1}{2}} = \left[\frac{R_L + iL\omega}{G + iC\omega}\right]^{\frac{1}{2}} . \tag{A1}$$

Assuming R_L and G to be negligible (the line will be distortionless), Z_K reduces to the pure resistance

$$R_K = (L/C)^{\frac{1}{2}} . \tag{A2}$$

If some line of arbitrary length l, terminated at one end in a load resistance R, is constructed such that $(L/C)^{\frac{1}{2}} = R$, there will be no reflection at the terminating end. Thus such a line is like an infinite line and has the characteristic impedance R which is a pure resistance. If such a line is thought of as a filter

it becomes apparent that the line will pass all frequencies because a pure resistance passess all frequencies. If this line is terminated at both ends in a resistance R, there will be no reflection at either end and all the power transmitted from one end will be received at the other end, provided the line is non-dissipative. Further, this matching of termination resistances is the condition for maximum transfer of power. With this as background, Nyquist's deduction of thermal noise equations can be readily understood, and we proceed to calculate the emf due to thermal noise.

Consider two noisy resistors each of value R and having uniform temperature T to be connected to each other by a long non-dissipative ideal transmission line as described immediately above. The equivalent circuit for this model is given in figure A1 where the noisy resistors are assumed to be representable by a voltage generator in series with R (noiseless).

Figure A1. Equivalent circuit of two noisy resistors, R and R, connected by a transmission line.

The noise voltage $\bar{v}$ generated in <u>one</u> of the resistors, say the one at the left, induces a current flow in the circuit whose value is given by $\bar{i} = \bar{v}/2R$. This current flowing through the other resistor causes it to heat up or absorb power, the absorbed power being given by $P = \overline{i^2}R = \overline{v^2}/4R$. That is, power is transferred from the one resistor to the other. The choice of resistor was arbitrary, and obviously there must be an identical transfer of power from the right resistor to the left, otherwise the temperature of each would not remain equal and unchanged, as required.

So far our ideal transmission line has provided the means for perfect--unrestricted--transfer of power between resistor. Now the line will also allow the means for reckoning the effect of frequency (bandwidth) on noise power and emf.

The line is non-reflecting and at any instant can be thought of as vibrating at its natural frequencies. If the length of line l is taken to be sufficiently large, the number of modes of vibration, or degrees of freedom, lying within a frequency range Δf is $2l\ \Delta f/v$, where v is the velocity of propagation. Then, the energy 'frozen' on the line at this instant is given by (the average energy per degree of freedom) X (the degrees of freedom). The average energy per degree of freedom is kT, where k is Boltzmann's constant. Thus the total energy of the vibrations within the frequency interval Δf is $kT2l\Delta f/v$.

But since there is no reflection, this is the energy within that frequency interval which was transferred from the two noisy resistors to the line during the transit time l/v.

Power is defined as energy per unit time. Thus the average power transferred from each resistor to the line within the frequency interval Δf during the time interval l/v is

$$P = \frac{1}{2}\ \frac{kT2l\Delta f/v}{l/v} = kT\Delta f\ . \tag{A3}$$

Thus the available thermal noise power--or maximum noise power which can be transferred by a resistor--is independent of the resistance value.

Earlier we derived from the equivalent circuit another relationship for the noise power, viz. $P = \overline{v^2}/4R$. Therefore,

$$\frac{\overline{v^2}}{4R} = kT\Delta f\ , \tag{A4}$$

and

$$\overline{v^2} = 4kTR\Delta f\ . \tag{A5}$$

Although the above equation was derived for R assumed to be pure resistance, similar reasoning shows it holds also for an impedance R(f) + X(f), when R in Eq. (A5) is repaced by R(f).

B. Derivation of Shot Noise

This derivation is made, as was the original, for a thermionic diode having temperature-limited, or saturated, emission (cf sec. 3.1.1). Nevertheless, the results hold equally well for any transport current which is by randomly produced free carriers; e.g., junction/barrier devices in general.

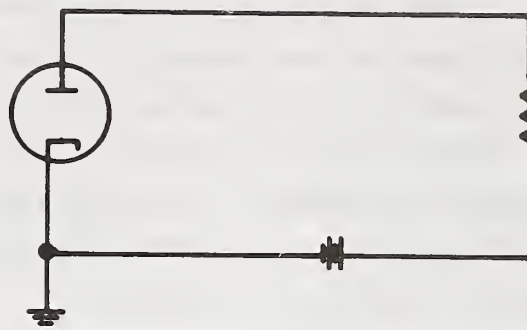

Figure B1. Thermionic diode circuit with plate resistor.

Figure B1 shows a thermionic diode circuit with plate (anode) resistor. The electrons emerge from the cathode randomly in time, and being free are distributed according to Maxwell's law with the minimum velocity just large enough to overcome the potential barrier at the surface (i.e., the work function); most electrons have velocities around the average value. The constant voltage applied to the anode establishes a constant electric field across the interspace between cathode and anode causing the emerging electron to be transported from cathode to anode with constant acceleration. The velocity of the electron is thus increased from its initial value v_i to some final value v_f, which is the velocity of the electron as it just reaches the anode. The effect of the electric field on velocity is very large so that $v_f >> v_i$. Thus the average velocity $\bar{v} \equiv (v_i + v_f)/2 \simeq v_f/2$.

Further, the applied voltage is sufficiently large so that the emission current is saturated. In saturated emission all emitted electrons reach the anode. Because the electrons are free the fluctuations in emission reach the anode unchanged. The random emergence of electrons from the cathode results in their random arrival at the anode.

The derivation of the shot noise equation is heuristically inspired in that the emission current arriving at the anode is perceived as a series of pulses. Owing to the corpuscular nature of the electron and the randomness in emission, the emission, and thus the plate current, varies from instant to instant. Measured over a long interval of time, however, the emission current averages out to an essentially constant value. Thus there is an instantaneous time-varying current fluctuating about an average, or direct current, in this case the saturated emission current.

Associated with the average current $\overline{I}$ are $\bar{n}$, the average number of electrons emitted per unit time and a constant; $\bar{v}$, the average velocity; and τ, the transit time, given by $\tau = d/\bar{v}$, where d is the cathode-to-anode spacing.

Associated with the instantaneous current I are n, the instantaneous rate of emission; $v(t)$, the instantaneous velocity; and t, the time at which the electron acquires a velocity v. A statistical theorem relates n and $\bar{n}$ (cf secs. 1.2 and 3.1.2), viz.:

$$\overline{(n-\bar{n})^2} = \bar{n} \ . \tag{B1}$$

The fluctuations in current $\overline{i}$ are defined by

$$\overline{i^2} \equiv \overline{(I-\overline{I})^2} \ . \tag{B2}$$

The derivation of the shot noise current equation--i.e., putting $\overline{i^2}$ into a practicable form--consists, in a nutshell, of 1) relating the average current and the instantaneous current at the plate by means of Campbell's (mathematical) theorem [31]; 2) introducing Fourier transforms to transform events from the time domain to the frequency domain, aided by Parseval's (mathematical) theorem [32]; 3) solving for $\overline{i^2}(f)$. Details of the mathematics are given below.

Applying Campbell's theorem:

$$\overline{I} = \bar{n} \int_{-\infty}^{\infty} F(t)\, dt \tag{B3}$$

and from eq. (B2)

$$\overline{i^2} = \bar{n} \int_{-\infty}^{\infty} F^2(t)\, dt \ . \tag{B4}$$

The Fourier transforms are now introduced:

$$F(t) = \frac{1}{(2\Pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} G(\omega) \exp(i\omega t)\, d\omega \,. \tag{B5}$$

$$G(\omega) = \frac{1}{(2\Pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} F(t) \exp(-i\omega t)\, dt \,, \tag{B6}$$

The Fourier transform of $F^2(t)$, obtained through use of Parseval's theorem, is

$$\int_{-\infty}^{\infty} F^2(t)\, dt = \int_{-\infty}^{\infty} G(\omega)\, G^*(\omega)\, d\omega \,, \tag{B7}$$

where $G^*(\omega)$ is the complex conjugate of $G(\omega)$. In the present example, $G(\omega)\ G^*(\omega)$ is an even function, so

$$\overline{i^2} = \frac{2I}{e} \int_{0}^{\infty} G(\omega)\, G^*(\omega)\, d\omega \,, \tag{B8}$$

where e is the electronic charge.

The function $F(t)$ is just the instantaneous current reaching the anode. The instantaneous current is given by

$$I = ev/d \,, \tag{B9}$$

where e is the electronic charge.

But $$d = v_f\, \tau/2 \,, \tag{B10}$$

and $$v = (t/\tau) v_f \,. \tag{B11}$$

Therefore, $$I = 2et/\tau^2 \,. \tag{B12}$$

Substituting this into eq. (B6) gives

$$G(\omega) = \frac{1}{(2\Pi)^{\frac{1}{2}}} \frac{2e}{\tau^2} \int_0^\tau t \exp(-i\omega t)\, dt \,. \tag{B13}$$

Denoting $\theta \equiv \omega\tau$ and integrating, gives

$$G(\omega) = \left(\frac{2}{\Pi}\right)^{\frac{1}{2}} \frac{e}{\theta^2} \left[(\theta \sin\theta + \cos\theta - 1) - i\,(\sin\theta - \theta\cos\theta)\right]. \tag{B14}$$

Then

$$G(\omega)\; G^*(\omega) = \frac{2e^2}{\Pi} \frac{1}{\theta^4} \left[\theta^2 + 2(1-\cos\theta-\theta\sin\theta)\right] . \tag{B15}$$

For saturated emission at low frequencies $\theta \approx 0$, and eq. (B15) reduces to

$$G(\omega)\; G^*(\omega) = e^2/2\Pi \,. \tag{B16}$$

Integrating eq. (B8) over the bandwidth Δf gives

$$\overline{i^2} = (eI/\Pi)\Delta\omega = 2eI\Delta f \,, \tag{B17}$$

which is the expression for the mean square shot noise, when I is the saturated current.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBS TN 1169	2. Performing Organ. Report No.	3. Publication Date December 1982

4. TITLE AND SUBTITLE

Introduction to Noise in Solid State Devices

5. AUTHOR(S)

Julius Cohen

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)	7. Contract/Grant No.
NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234	8. Type of Report & Period Covered Final

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

Same as item 6.

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

This is a short didactic monograph on electronic noise which aims to impart a "feel for the subject". The work is divided into two parts. The first, Theory, deals in detail with the principal noises found in solid state devices, namely shot noise, thermal noise, 1/f noise, and generation-recombination noise. The second part, Applications, is a systematic treatment of noise in selected solid state devices. Analyses progress from a single noise source in a circuit element to four noises in a device; concomittantly equivalent circuits are developed to facilitate the solution of various complex noise problems. Examples treated in this part include resistors, rectifiers, transistors, and photo-detectors. The work concludes with a recapitulation and useful references.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

electronics; noise; photon detector; rectifier; solid state devices; transistor.

13. AVAILABILITY	14. NO. OF PRINTED PAGES
[X] Unlimited ☐ For Official Distribution. Do Not Release to NTIS [X] Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. ☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161	63 **15. Price** $4.75

USCOMM-DC 6043-P80

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent Bureau publications in both NBS and non-NBS media. Issued six times a year. Annual subscription: domestic $18; foreign $22.50. Single copy, $4.25 domestic; $5.35 foreign.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NBS under the authority of the National Standard Data Act (Public Law 90-396).

NOTE: The principal publication outlet for the foregoing data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the **following** *NBS publications—FIPS and NBSIR's—from the National Technical Information Services, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Services, Springfield, VA 22161, in paper copy or microfiche form.